

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2003, 16:34:19 ; Search time 6488 Seconds
(without alignments)
16982.618 Million cell updates/sec

Title: US-09-715-876-7
Perfect score: 3786
Sequence: 1 atgcttcaacaattacatt.....tgtctgttcattagtga 3786

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_da.*
- 2: gb_hgt.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_or.*
- 21: em_ov.*
- 22: em_pat.*
- 23: em_ph.*
- 24: em_pl.*
- 25: em_ro.*
- 26: em_sts.*
- 27: em_un.*
- 28: em_vi.*
- 29: em_hgt_hum.*
- 30: em_hgt_inv.*
- 31: em_hgt_mus.*
- 32: em_hgt_other.*
- 33: em_hgt_pln.*
- 34: em_hgt_rdt.*
- 35: em_hgt_rod.*
- 36: em_hgt_nam.*
- 37: em_hgt_vrt.*
- 38: em_sy.*
- 39: em_hgtgo_hum.*
- 40: em_hgtgo_mus.*
- 41: em_hgtgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3786	100.0	3786	8	YSAALS1	L25902 Candida alb
2	1834.4	48.5	3360	8	CAU87956	U87956 Candida alb
3	1708.6	45.1	4383	8	AF051313	AF051313 Candida a
4	1526.4	40.3	4569	8	AF272027	AF272027 Candida a
5	1392.8	36.8	4723	8	AF025429	AF025429 Candida a
6	1188.8	31.4	3813	8	AF068866	AF068866 Candida a
7	1069.4	28.2	1071	6	AR044076	AR044076 Sequence 1
8	1069.4	28.2	1071	6	165632	165632 Sequence 1
9	1016.6	26.9	1404	8	CANALS2S1	AF024582 Candida a
10	1006	26.6	1404	8	CAIALS1	AF024580 Candida a
11	905.8	23.9	4332	8	AF075293	AF075293 Candida a
12	874.4	23.1	6897	8	AF201684	AF201684 Candida a
13	842.8	22.3	1407	8	CANALS4S1	AF024584 Candida a
14	836.4	22.1	1407	8	CANALS4S1	AF024586 Candida a
15	805.4	21.3	1404	8	AF229989S1	AF229989 Candida a
16	731.8	19.3	1007	8	AF202529	AF202529 Candida d
17	507	13.4	995	8	AF202530	AF202530 Candida d
18	463.6	12.2	1013	8	AF201685	AF201685 Candida d
19	394	10.4	1299	8	AF075294	AF075294 Candida a
20	362.2	9.6	1058	8	AF201686	AF201686 Candida t
21	348.2	9.2	425	6	AR044080	AR044080 Sequence 5
22	348.2	9.2	425	6	165636	165636 Sequence 5
23	331.2	8.7	680	8	AF189016	AF189016 Candida a
24	202.4	5.3	1400	8	AB002099	AB002099 Candida t
25	198.6	5.2	13370	8	SPAPB2C8	AL590602 S.pombe c
26	170.8	4.5	172307	2	AC044842	AC044842 Homo sapi
27	141.6	3.7	377	8	AF211866	AF211866 Candida t
28	135.8	3.6	132449	9	AL365272	AL365272 Human DNA
29	134.4	3.5	20325	8	SPCPB16A4	AL591677 S.pombe c
30	132	3.5	380	8	AF211865	AF211865 Candida t
31	130.4	3.4	158615	2	AC117835	AC117835 Rattus no
32	126	3.4	185994	2	AC002042	AC002042 Homo sapi
33	125.8	3.3	175748	2	AC120669	AC120669 Rattus no
34	125.8	3.3	180903	2	AC125859	AC125859 Rattus no
35	121.6	3.2	12151	1	U23947	U23947 Mycoplasma
36	116	3.1	53352	5	AL592077	AL592077 Zebrafish
37	112.2	3.0	157	8	AF035757	AF035757 Candida a
38	111.8	3.0	180668	2	AC020857	AC020857 Mus muscu
39	111.6	2.9	666	8	AF413051	AF413051 Zea mays
40	111.2	2.9	22398	5	FR0271723	AF271723 Fugu rubr
41	110.2	2.9	141017	2	AC116962	AC116962 Dictyoste
42	109.8	2.9	35412	8	SPCC188	AL049662 S.pombe c
43	109.8	2.9	38141	8	SPAC8A4	266369 S.pombe chr
44	109.6	2.9	127354	2	AC117014	AC117014 Rattus no
45	109.6	2.9	155019	2	AC117361	AC117361 Rattus no

ALIGNMENTS

RESULT 1	YSAALS1	Candida albicans agglutinin-like sequence (ALS1) gene, complete cds.	3786 bp	DNA	linear	PLN 03-MAY-2000
LOCUS DEFINITION	YSAALS1	Candida albicans agglutinin-like sequence (ALS1) gene, complete cds.				
ACCESSION	L25902					
VERSION	L25902.1	GI:704426				
KEYWORDS						
SOURCE		Candida albicans				
ORGANISM		Candida albicans				
REFERENCE		Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.				
AUTHORS		1 (bases 1 to 3786)				
TITLE		Hoyer, L.L., Scherer, S., Shatzman, A.R. and Llivi, G.P.				
		Candida albicans ALS1: domains related to a Saccharomycetes				

cerevisiae sexual agglutinin separated by a repeating motif

Mol. Microbiol. 15 (1), 39-54 (1995)

95272392

7752895

FEATURES
source

source

Location/Qualifiers

1. .3786

/organism="Candida albicans"

/strain="B792"

```
/db xref="taxon:5476"
```

700-3783

```
1. 3703
/ gene="AT.S1"
```

1 3783

1. 3/83
/CONE # AT.51 #

/gene= ATSI
/codon atatat-1

```
/usr/local/src/openssl-1.0.2g/openssl/cod.../codon_start=1
```

```
/transi_table=12
```

/product="agglutinin-like

/protein_id="AAC41649.2"

```

/db_xref="MQOQTLFLFLYSIASAKTITGVDFSNLSWNSAANYAFKGP
/translation="MLQOQTLFLFLYSIASAKTITGVDFSNLSWNSAANYAFKGP
EPTTNALCISLDGTSANPGDFTILNMPGVFYVTSQTSDDTDAGVKVATQCFNVSE
YFTTSTVINDKALKSIIKAGVTWLPFAENVGSGTSDLEDKCFATACQNTVTF
NNGDKXISIDVEFEKSTVDPXAYLYASRVMPSLNKVTLFLVAPQCENGYTSGMTGFS
SNGDVAISNHIHIGLKNLDWNPVSSSEFYTKTCSNGIQIKYQNPVAGYRPFI
DAYISATDYNQVTLAYINDCAGSSQSKPFLRWTKYKNSDAQISQIVVATRTV
TDSSTVATLTPNPSYDKTKTEILQIPTTTITTSYGVGTSYTKTAPIGETATVI
VDVPHVHTTITVSEMTGTTTTTTRTNPDTSDIVVQVSPNPTVSTTEYWSQSFAT
VITVAPPGTGDVIIIREPNNHTVITEYWSQSFATTTVATPAGGTDSVIIIREPNN
TVTTEYWSQSFATTTVAPPGTGDVIIIREPNNPTVTTEYWSQSYATTTVATP
GGTDSVIIIREPNNHTVITEYWSQSYATTTVAPPGTGDVIIIREPNNPTVTTEY
SOFATTTVTPPGSGTDVIIIREPNNPTVTTEYWSQSYATTTITAPPGTGDVII
REPPNNVTTEYWSQSYATTTVAPPGTGDVIIIREPNNHTVITEYWSQSYATTT
TVAPPGTGDVIIIREPNNPTVTTEYWSQSFATTTVAPPGTGDVIIIRESMSSK
ISTSDNIIIPSEPRHYVNSTSDLTFESSMNTPTSISSDGLMLSSITLVTES
ITFTESICDGKECSRSLSSGIGVTPNDNESSIVTSTVPTASTMSDSLSTGDISAT
SDNWKSGVSTTEITVTYTIOTFNPLSSVTSLTQLSIPSESESKVFTPSNGD
NOSGHDGSGSTEBIIVTTSIKVLPVNSNDLTSEPTNRPQPTLTSTNSIT
EDITTSQPGDNGDNTSNPVTVAITSFLASEEDNKSGSHESASTSLKPSMGENS
GLUTTEIATEATLSTPEASPAVSSGDTVTTEPTREQTTLSTTSKTNSEVATQ
ATNENKSGPSDLTSLTGTSTANSSELVTSIGSVTGGAVASANDQSHSTSVTN
SNSIVSNTPTTLTSQOVSSTSPSTFIASTDVGSGSIIHGTWLYGLTLTLLSLFT"

```

Query Match

Query Method	Best Local Similarity	100.0%	Score	prod No
Query Method	Best Local Similarity	100.0%	Score	prod No

BEST LOCAL SIMILARITY 100.0%, FREQ. NO. 0;
Matches 3796, Conservative 0, Mismatches 0;
Indels 0, Cons 0;

	Qy	1	ATGCTTCACAACAAATTACATGTGTTATTCCTATATTTGGTCAATTCGAAGTGCAAGAACAATC	60
	Db	1	ATGCTTCAACAACAAATTACATGTGTTATTCCTATATTTGGTCAATTCGAAGTGCAAGAACAATC	60
	Qy	61	ACTGGTGTTTTGGATAGTTTTAAATTCATTAACTTGGTCCAATGCCTGAATTAATATGCTTTTC	120
	Db	61	ACTGGTGTTTTGATAGTTTTAAATTCATTAACTTGGTCCAATGCCTGAATTAATATGCTTTTC	120
	Qy	121	AAGGCCCAAGGATACCACACTTGGGAATGCTGTTTTGGGTGGTCCCTAGATGGTACCAGT	180
	Db	121	AAGGCCCAAGGATACCACACTTGGGAATGCTGTTTTGGGTGGTCCCTAGATGGTACCAGT	180
	Qy	181	GCCAACTCCAGGGGATACATTTCATTTGAATATGCCATGTGCTGTTTAAATATACTACTTCA	240
	Db	181	GCCAACTCCAGGGGATACATTTCATTTGAATATGCCATGTGCTGTTTAAATATACTACTTCA	240
	Qy	241	CRAACATCTGTTGATTAACTCGCGATGGTGTTAAATATGCTACTTGTCATTTTATTCT	300
	Db	241	CRAACATCTGTTGATTAACTCGCGATGGTGTTAAATATGCTACTTGTCATTTTATTCT	300
	Qy	301	GGTGAAGAANTCACAACTTTTCTCATATTAACATGATCTGTGAACGACGCTTTGAAATCA	360
	Db	301	GGTGAAGAANTCACAACTTTTCTCATATTAACATGATCTGTGAACGACGCTTTGAAATCA	360

|||||
Db 1441 TCACAATCCTTGTCTACTACTACTACTACTACTCTCTCCAGTGGTACTGACTCAGTA 1500
QY 1501 ATTATCAGAGAACCACCAATCAACTGTCTACTACACCGAGTATGGTCTCAATCCTTT 1560
Db 1501 ATTATCAGAGAACCACCAATCAACTGTCTACTACACCGAGTATGGTCTCAATCCTTT 1560
QY 1561 GCTACTACTACTACAGTTACTGCTCTCCAGGTGGTACTGACTCACTAATATATCAGAGAA 1620
Db 1561 GCTACTACTACTACAGTTACTGCTCTCCAGGTGGTACTGACTCACTAATATATCAGAGAA 1620
QY 1621 CTTCCAAACCCAACTGTACACCACTGAATATATGGTCCCAATCTTTACGCAACCAACT 1680
Db 1621 CTTCCAAACCCAACTGTACACCACTGAATATATGGTCCCAATCTTTACGCAACCAACT 1680
QY 1681 ACTGTGACTGCTCTCCAGGAGCACTGACTCAGTCAATATATCAGAGAACCCACCAACCAC 1740
Db 1681 ACTGTGACTGCTCTCCAGGAGCACTGACTCAGTCAATATATCAGAGAACCCACCAACCAC 1740
QY 1741 ACTGTGACTACTACTGAATACTGGTGCACAAATCATATGCCACCACTACCACCTGTAACCTGCA 1800
Db 1741 ACTGTGACTACTACTGAATACTGGTGCACAAATCATATGCCACCACTACCACCTGTAACCTGCA 1800
QY 1801 CCACGAGGTGGTACTGACACTGTTATCATATAGAGAGCCACCAACCACTGTCACTACT 1860
Db 1801 CCACGAGGTGGTACTGACACTGTTATCATATAGAGAGCCACCAACCACTGTCACTACT 1860
QY 1861 ACTGAGTATGGTCTCAATGCTTGTCTACTPACCACAACTGTAACCTGGTCCACCAAGTGGC 1920
Db 1861 ACTGAGTATGGTCTCAATGCTTGTCTACTPACCACAACTGTAACCTGGTCCACCAAGTGGC 1920
QY 1921 ACTGATACTGTTATCATATAGGGAACCAACCAACCCAACTGTCAACCACTACTGTAATCTGG 1980
Db 1921 ACTGATACTGTTATCATATAGGGAACCAACCAACCCAACTGTCAACCACTACTGTAATCTGG 1980
QY 1981 TCTCAATCATATGCAACCACTACTACCATACCGCTCCACCTGGTGAACCTGATACCGTT 2040
Db 1981 TCTCAATCATATGCAACCACTACTACCATACCGCTCCACCTGGTGAACCTGATACCGTT 2040
QY 2041 CTTATCAGAGAGCCACCAACCACTACTGTCACTACTACTGAATCTGGTCTCAATCATAT 2100
Db 2041 CTTATCAGAGAGCCACCAACCACTACTGTCACTACTACTGAATCTGGTCTCAATCATAT 2100
QY 2101 GCTACAAACCACTGTTACTGCACCACTGGTGAACCGATACCGTCTTATCAGAGAG 2160
Db 2101 GCTACAAACCACTGTTACTGCACCACTGGTGAACCGATACCGTCTTATCAGAGAG 2160
QY 2161 CCACCAACCACTACTGCTACTACTGAATCTGGTCTCAATCATATGCTACAAACCACT 2220
Db 2161 CCACCAACCACTACTGCTACTACTGAATCTGGTCTCAATCATATGCTACAAACCACT 2220
QY 2221 ACTGTTACTGCACCACTGGTACCGATCTGTTATCATATAGAGAGCCACCAATPCCA 2280
Db 2221 ACTGTTACTGCACCACTGGTACCGATCTGTTATCATATAGAGAGCCACCAATPCCA 2280
QY 2281 ACAGTTACTACTACTGAATATGGTGCACAAATCATATTTGCCACCAACCACTGTTACTGCT 2340
Db 2281 ACAGTTACTACTACTGAATATGGTGCACAAATCATATTTGCCACCAACCACTGTTACTGCT 2340
QY 2341 CTTCCAGGTGGTACTGACACTGTTATCTATGAAAGCATGTCAAGTTCAAGATTTCT 2400
Db 2341 CTTCCAGGTGGTACTGACACTGTTATCTATGAAAGCATGTCAAGTTCAAGATTTCT 2400
QY 2401 ACATCTCCAAATGATATAACCACTATCATTTCCCGTCTCTCATATATGTCAAC 2460
Db 2401 ACATCTCCAAATGATATAACCACTATCATTTCCCGTCTCTCATATATGTCAAC 2460
QY 2461 AGCACAACCTCCGATTTGTCAACATTTGAATCTTCAATCCCAATGATCTCTCTCTATC 2520
Db 2461 AGCACAACCTCCGATTTGTCAACATTTGAATCTTCAATCCCAATGATCTCTCTCTATC 2520
QY 2521 AGTAGTAGGTATGTTGTTCTTCTACAACTTTGGTACTGAATCAGAGAAACCACTACA 2580
|||||

Db 2521 AGTAGTAGGTATGTTGTTCTTCTACAACTTTGGTACTGAATCAGAGAAACCACTACA 2580
QY 2581 GAACTGATTTGCAGTGTATGTAAGAGTGTCTAGATTTGTCAGTGTCTCTGCTATTGTC 2640
Db 2581 GAACTGATTTGCAGTGTATGTAAGAGTGTCTAGATTTGTCAGTGTCTCTGCTATTGTC 2640
QY 2641 ACAAAATCCAGATAGCAATGAATCCTCAATGTAAGTACTGTCTCTACTGCAAGTACA 2700
Db 2641 ACAAAATCCAGATAGCAATGAATCCTCAATGTAAGTACTGTCTCTACTGCAAGTACA 2700
QY 2701 ATGCTGATTTCACTTTCTTCAACTGATGTTAGTGTCTGCTACTGATTAATGTTTCA 2760
Db 2701 ATGCTGATTTCACTTTCTTCAACTGATGTTAGTGTCTGCTACTGATTAATGTTTCA 2760
QY 2761 AAATCAGGAGTATCAGTTACAAACCGAAACTTCTGTTACAACTATTCAAACTACTCCAAAC 2820
Db 2761 AAATCAGGAGTATCAGTTACAAACCGAAACTTCTGTTACAACTATTCAAACTACTCCAAAC 2820
QY 2821 CCATTTATCATCTTCTCAGTGACATCATTTGACTGACTGTCTTCAATTTCAAGTGTTCAGAA 2880
Db 2821 CCATTTATCATCTTCTCAGTGACATCATTTGACTGACTGTCTTCAATTTCAAGTGTTCAGAA 2880
QY 2881 AGTGAAGTAAAGTTACATTTACAAGCAATGGAGACAACCAAGTGGTACTGATTTCA 2940
Db 2881 AGTGAAGTAAAGTTACATTTACAAGCAATGGAGACAACCAAGTGGTACTGATTTCA 2940
QY 2941 CAATCTACTTCCACTGAAATTTGAAATTTGTAACCAACAGTGTCTACTAAAGTGTTCACCACT 3000
Db 2941 CAATCTACTTCCACTGAAATTTGAAATTTGTAACCAACAGTGTCTACTAAAGTGTTCACCACT 3000
QY 3001 GTCGTTTCTTCTAATCTGATTTGACTGTAGTGAACCAACCAATACCAGAGAACCAACCACT 3060
Db 3001 GTCGTTTCTTCTAATCTGATTTGACTGTAGTGAACCAACCAATACCAGAGAACCAACCACT 3060
QY 3061 ACATTTACCACTACTTCAAACTCCATCCTGAAAGATATACCACATCTCAACCTACAGGT 3120
Db 3061 ACATTTACCACTACTTCAAACTCCATCCTGAAAGATATACCACATCTCAACCTACAGGT 3120
QY 3121 GATAATGGAGACAATCTTCAATCAACCAATCCAGTTCCTCAATGAAATCAGCATCCACAAGTTTGA 3180
Db 3121 GATAATGGAGACAATCTTCAATCAACCAATCCAGTTCCTCAATGAAATCAGCATCCACAAGTTTGA 3180
QY 3181 GCATCTGCAAGTGAAGAGACAACAAAGCGGTTCTCATGAAATCAGCATCCACAAGTTTGA 3240
Db 3181 GCATCTGCAAGTGAAGAGACAACAAAGCGGTTCTCATGAAATCAGCATCCACAAGTTTGA 3240
QY 3241 AAACCAAGTATGGTGAAGATTTCTGGATTAACCTACTTCTACTGAAATTTGAAGCTTACAACA 3300
Db 3241 AAACCAAGTATGGTGAAGATTTCTGGATTAACCTACTTCTACTGAAATTTGAAGCTTACAACA 3300
QY 3301 ACCAGTCTACAGAAGCTCCATCACCCTGCTGTTTCTTCTGCTACTGATGTAACCTACTGAA 3360
Db 3301 ACCAGTCTACAGAAGCTCCATCACCCTGCTGTTTCTTCTGCTACTGATGTAACCTACTGAA 3360
QY 3361 CCAACTGATACTAGAGAACCACTACTACATTTATCAACTACTTCAAAAACCAACAGTGA 3420
Db 3361 CCAACTGATACTAGAGAACCACTACTACATTTATCAACTACTTCAAAAACCAACAGTGA 3420
QY 3421 CTGTTGCTACTACACAAGCTACTAATGAAATGGTGGTAAATCTCCATCAACTGATTTTA 3480
Db 3421 CTGTTGCTACTACACAAGCTACTAATGAAATGGTGGTAAATCTCCATCAACTGATTTTA 3480
QY 3481 ACATCAAGCTTTGACACAGGCACTCAGCATCTACAAGTGTCTAATAGCGAAGTGTGTACT 3540
Db 3481 ACATCAAGCTTTGACACAGGCACTCAGCATCTACAAGTGTCTAATAGCGAAGTGTGTACT 3540
QY 3541 AGTGGATCTGTTACTGGTGGAGCTGTTGCCAGTGTTCAAATGATCAATCAGATCTTACT 3600
Db 3541 AGTGGATCTGTTACTGGTGGAGCTGTTGCCAGTGTTCAAATGATCAATCAGATCTTACT 3600
QY 3601 TCTGTTTACCAACAGCACAGCATTTGATCTAATACCCCAACCAACTACATTTGAGTCAACAA 3660
Db 3601 TCTGTTTACCAACAGCACAGCATTTGATCTAATACCCCAACCAACTACATTTGAGTCAACAA 3660

```
Qy 3661 GTTACCTCATCTCACCCTTCAACCAACACATTCATTCGCTTCTACATACGATGGCTCTGGT 3720
|||||
Db 3661 GTTACCTCATCTCACCCTTCAACCAACACATTCATTCGCTTCTACATACGATGGCTCTGGT 3720
|||||
Qy 3721 TCTATTCCAAACATCTTACTTGGTTGACGGTTTGATCACATTTATGTCCTTGTTCATT 3780
|||||
Db 3721 TCTATTCCAAACATCTTACTTGGTTGACGGTTTGATCACATTTATGTCCTTGTTCATT 3780
|||||
Qy 3781 TAGTGA 3786
|||||
Db 3781 TAGTGA 3786

RESULT 2
CAU87956 3360 bp DNA linear PLN 02-JUL-1998
LOCUS Candida albicans agglutinin-like protein (ALS3) gene, complete cds.
DEFINITION U87956
ACCESSION U87956
VERSION U87956.1 GI:3273414
KEYWORDS
SOURCE Candida albicans.
ORGANISM Candida albicans.
Eukaryota: Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; microsporid Saccharomycetales; Candida.
REFERENCE
AUTHORS Hoyer,L.L., Payne,T.L., Bell,M., Myers,A.M. and Scherer,S.
TITLE Candida albicans ALS3 and insights into the nature of the ALS gene
family
JOURNAL Curr. Genet. 33 (6), 451-459 (1998)
MEDLINE 98309840
PUBMED 9644209
REFERENCE
AUTHORS Hoyer,L.L.
TITLE Direct Submission
JOURNAL Submitted (30-JAN-1997) Veterinary Pathobiology, University of
Illinois, 2001 S. Lincoln Avenue, Urbana, IL 61802, USA
FEATURES
source
1. .3360
/organism="Candida albicans"
/strain="1161"
/db.xref="taxon:5476"
/chromosome="R"
/map="5+0"
1. .3360
/gene="ALS3"
1. .3360
/gene="ALS3"
/codon_start=1
/trans_table=12
/product="agglutinin-like protein"
/protein_id="AAC39486.1"
/db.xref="GI:3273415"
/translations="MLQOYVLLIYLVSATAKITGVNSFNSLTWSNATYNYKPGG
TPYNAVLSHSLDGTASPGDTEFLANPCVFKETTSQTSVDLTAHGKATYACQFQAGE
EFMTFSLTCTVNTLPSKALGTVLPLAFNVGGTSSVDLESKCFAGTNTVTF
NDGKQISINDIERNVDPKGLTDSRVIPSLNKVSTLFVAPCCANGYTSGMTGFAN
TYGDVQIDNIHVHIGITKGLDWNYPVSEFSYTKTSSNGIFITYKNVPAGYRPFV
DAISATDVNSYVLSYANEXYTCAGYQWAPFTLRWTGYRNSDAGSNGIVIVATRTV
TDSATVTLTPDPNRDKTKTIEILKPIPTTTTTSVGVTTSTYKTAPIGTATVI
VDIPIHTVTTTSKWTCTPSNTHTNPDSIDTVIVQVSPNPTVTTTEYWSQSEAT
TTVITGPGNVDVLIREPNNHVTTEYWSQSEATTTSTFTAPPGSDVSLIREFPNP
TVITTEYWSQSEATTTSTFTAPPGSDVSLIREFPNHVTTEYWSQSEATTTTVPAPP
GGDVLVREPNNHVTTEYWSQSEATTTTVPAPPDSDVLIREFPNHVTTEYWSQ
SQYATTTTTPAGGETDVLIREPNNHVTTEYWSQSEATTTTTPAGGETDVLIREP
REPNHVTTEYWSQSEATTTTVPAGGETDVLIREPNNHVTTEYWSQSEATTTT
TTPAGGETDVLIREPNNHVTTEYWSQSEATTTTTPAGGETDVLIREPNNHVT
TTEYWSQSEATTTTVPAGGETDVLIREPNNHVTTEYWSQSEATTTTTPAGGETD
ETKTTTSCGGKGSWSVSRIVTIPNNIETPMVNTVDSPTTSTESTSQSPSGIFS
ESGVSESTESTVTTAQTPNSVPTTESEVFTTKGNENGPYESPTNKKSSDENSE
FTTSTAESTDENETIATGVSSEASSPISSADETTTVAESTSVIEQPTNNN
GGKAPASATSSPTTTTANNDSVITGTSTNQSSQSQYNSDFTQQTTLSSQMTSSLVS
LHMLTFDGSQSVIQTSLWLCGLITLLSLFI"
```

```
BASE COUNT 1029 a 804 c 520 g 1007 t
ORIGIN
Query Match 48.5%; Score 1834.4; DB 8; Length 3360;
Best Local Similarity 73.9%; Pred. No. 0;
Matches 2393; Conservative 0; Mismatches 830; Indels 15; Gaps 5;
Qy 1 ATGCTTCAACAATTTACATTCGTTATTCCTATATTTGTCAATTTGCAAGTGCACAAACATC 60
|||||
Db 1 ATGCTACAACAATATACATTCGTTACTCATATATTTGCGTTGGCACTGCACAAACATC 60
|||||
Qy 61 ACTGGTGTGTTTGTAGTGTGTTTAAATTCATTAACTTGGTCCAATGCTGCTAAATTCGTTTC 120
|||||
Db 61 ACTGGTGTGTTTCAACAGTGTGTTTAAATTCATTAACTTGGTCCAATGCTGCTAAATTCATTAATAT 120
|||||
Qy 121 AAAGGGCCAGATACCAACTTTGAATGCTGCTGTTGGTGGTCTTGTAGATGGTACCAGT 180
|||||
Db 121 AAGGGACAGGAACCCCAACTTTGAATGCTGCTGTTGGTGGTCTTGTAGATGGTACCAGT 180
|||||
Qy 181 GCCAATCCAGGGATACATTCACATTTGAATGCTGCTGTTGTTTAAATATACATCTCA 240
|||||
Db 181 GCAAGTCCGGGAGATACATTCACATTTGAATGCTGCTGTTGTTTAAATATACATCTCT 240
|||||
Qy 241 CAACATCTGTGTTGATTTAACTGCCGATGGTGTAAATATGCTACTTGTCAATTTTATCT 300
|||||
Db 241 CAACATCTGTGTTGATTTGACTGCTCATGGTGTAAATATGCTACTTGTCAATTTTCAAGCA 300
|||||
Qy 301 GGTCAAGAATTTACCAACTTTTCTACATTAACATGTTACTGTGAACGCGCTTTGAAATCA 360
|||||
Db 301 GGTCAAGAATTTAGCACTTTTCTACATTAACATGTTACTGTGAACGCAATACATTTGACTCCA 360
|||||
Qy 361 TCCATTAAAGCATTTGGTACAGTTACTTTACCAATTCATTCATTTGTTGGTGGCAAGGT 420
|||||
Db 361 TCTATTAAAGCTTTGGGTACTGTCCACTTACCCTTGCATTCATTTAGTGGTGAACGTGT 420
|||||
Qy 421 TCATCACTGATTTGGGAGATTTCTAAATGTTTACTGCTGTGTACCAATACATGATCAGTTCACATTT 480
|||||
Db 421 TCTTCTGTTGATTTGGAAGATTTCTAAATGTTTACTGCTGTGTACCAATACATGATCAGTTCACATTT 480
|||||
Qy 481 AATGATGCTGATAAGATATCTCAATTCATTCAGTTTGTGAGTTTGAAGATCAACGTTTGATCCA 540
|||||
Db 481 AATGATGTCGCAAGAAAATCTCTATTAAAGTTGATTTGAAAGTCAATATGTCGATCCA 540
|||||
Qy 541 AGTCATATTTGTATGCTCCAGAGTTATGCCAAGTCTCAATAAGGTTCACAACTCTTTT 600
|||||
Db 541 AAAGGTACTTAAGTATCCAGAGTTATCCAGAGTTATCCAAAGTCTCAACAAAGTCAACCTCTTTT 600
|||||
Qy 601 GTGCACACCAATCTGAAATGGTTACACATCTCGTACATCTCGTACATGGGGTCTCCAGTAGTAAC 660
|||||
Db 601 GTTCACACCAATCTGCAATGGTTACACATCTCGTACATGGGGTCTCGCTTAACACTAT 660
|||||
Qy 661 GGTGACGTTGCTATTGATTCGTTCAATATTCATATTGTTGATCACAAGAGGATTAATATGAT 720
|||||
Db 661 GGTGATGTTCAATGACTGTTCAATATTCATGTTGTTGATTAACAAGGATTAATATGAT 720
|||||
Qy 721 TGAATATCCGGTTCATCTGGAATCATTTAGTTACACTAAAACCTGTACATCTAATGGA 780
|||||
Db 721 TGAATATCCGGTTCATCTGGAATCATTTAGTTACACTAAAACCTGTCTTAATGTTT 780
|||||
Qy 781 ATTACAGATTAATATCAAAATGACCTGCTGGTTATCGTCCATTTATTTGATGCTTATATT 840
|||||
Db 781 ATCTTTATCAATATAAAATGTTCTCGCCGGTTATCGTCCATTTGTTGACGCTTATATT 840
|||||
Qy 841 TCTGCTACAGATTTAAACCAATATCTTTAGCATATACCAATGATTAATATCTGCTGGC 900
|||||
Db 841 TCTGCTACAGATTTAAATTCGTACACCTTGTGCTGATGCTAATGATATCTGCTGGT 900
|||||
Qy 901 AGTCGCTGCAAACTAAACCTTTCACTTAAAGATGGACTGGATACAGAAATAGTATGCC 960
|||||
Db 901 GGTATTATGGCAACGTGACCTTTTCAATATAGATGGACTGGATACAGAAATAGTATGCC 960
|||||
Qy 961 GGATCTAACGGTATTGTCATTTGTTGCTACACTAGACAGTTACAGACAGTACCACCTGCT 1020
|||||
```

Db	961	GGATCTAACGGTATNTGTTATTTGCGCTACTACCAAGACAGTTACAGACAGTACTACCGCT	1020
Qy	1021	GTCACTACTTTTACCAATCCAAAGTGTGATAAAAACCAAAACAAATCGAAATTTTGGCA	1080
Db	1021	GTGACCACCTTACCATTCGATCCTTAACCGCGACAAACTAGACAATTGAATTTTGAA	1080
Qy	1081	CCTATTCCAAACACATACCAATCAACACTTCATATGTTGGTGTGACTACTTCTCTATCTGACT	1140
Db	1081	CCTATTCCAAACAACTACAATCAACAACATCATATGTTGGTGTGACTACTTCTCTACCTGACC	1140
Qy	1141	AAGACTGCACCAATTTGGTGGAAACAGCTACTGTTATTTGTTGATGTGCCATATCATACTACC	1200
Db	1141	AAACTGCAACCAATTTGGGGAACTGCTACTGTTATTTGTTGATATTCATATCAACACTACC	1200
Qy	1201	ACAACCTGTTACCAAGTAATGAGACAGAAACAATCACTACCAACCAACTCGTACCAATCCA	1260
Db	1201	ACTACTGTTACCAAGTAATTTGGACAGAAACAATTACTTCCACCAACACATACTAATCCA	1260
Qy	1261	ACTGATTCAATTGACACAGTGGTGTGACAAGTTCCACTGCGCAAAATCCAACCTGTGTAGTACT	1320
Db	1261	ACTGACTCAATAGACACTGCTATTTGTACAAGTTCCACTGCGCAAAACCAACTGTTTACTACC	1320
Qy	1321	ACTGAATATTGGTCTCAGTCTTTCCTCAACCACTACAGTTACTGCTCTCCAGGTGGT	1380
Db	1321	ACTGAATATTGGTCTCAATCATTTTTTCTACCAACCAACCACTTACTGACCACCAAGAAAC	1380
Qy	1381	ACCGATACTGTGATTAATCAGACAGCCACAAACCAATACGTGCACTACTACTGTAATATTGG	1440
Db	1381	ACTGATACTGTTTTAATCAGACAGCCAGCAACCACTACTGTCACTCAACCGAGTACTGG	1440
Qy	1441	TCACAATCCTTTGCTACTACTACTACTGTTACTGCTCTCCAGGTGGTACTGACTCAGTA	1500
Db	1441	TCAGAATCTTTACACTACTACTAGTACTTTCACTGCTCTCCAGGTGGAACCTGATTCAGTT	1500
Qy	1501	ATTATCAGAGAACCAACCAATCCAACTGTCACTCAACCGAGTATTGGTCTCAATCCTTT	1560
Db	1501	ATTATCAAGGAACCTCCAAATCCAACTGTCAACACTACCGAGTACTGGTCAGAACTTTAC	1560
Qy	1561	GCTACTACTACTACAGTTTACTGCTCTCCAGGTGGTACTGACTCAGTAATTAATCAGAGAA	1620
Db	1561	ACTACCAGTAGTACCTTCACTGCTCTCCAGGTGGAACCTGATTCAATTAATCAAGGA	1620
Qy	1621	CCTCCAAACCCAACTGTCAACCACTGAATATTGGTCCCAATCTTAGCAACCAACACT	1680
Db	1621	CCACCAAAACCATACTGTAAACCAACACTGAATATTGGTCACAATCTTACACTACCACACT	1680
Qy	1681	ACTGTGACTGCTCCTCCAGSAGSACTGACTCAGTAATTAATCAGAGAACCAACCAACCAC	1740
Db	1681	ACTGTCAACGCTCCACAGGAGTATGATACTGTCTTAGTCAGAGAACCAACCAACCAT	1740
Qy	1741	ACTGTCACTACTACTGAATACTGGTCACAATCATATGCCACACTACCACTGTAACTGCA	1800
Db	1741	ACTGTTACAACTACCGAGTACTTGGTCACAACTTACACTCAACCAACCATGTTATTGGC	1800
Qy	1801	CCACCAAGTGGTACTGACACTGTTATCATTTAGAGAGCCACCAACCACTGTCTACTACT	1860
Db	1801	CCACCAAGTGGCACTGATTCGGTTATCATTTAGAGAAACCTCCAAATCCAACCTGTCAACA	1860
Qy	1861	ACTGAGTATTGGTCTCAATCGTTTGGTACTACCACAAGTGAATGGTCCACCAAGTGGC	1920
Db	1861	ACTGAGTACTGGTCTCAATCTTACGCACTACCACTACCATTAACCGGTCTCTCCAGGTGAG	1920
Qy	1921	ACTGATACTGTTATCATTTAGGGNACCAACCAACCACTGTCCACTACTGAATACTGG	1980
Db	1921	ACCGATACTGCTCTTATTAGAGAACCAACCAACCACTGTGAACCAACCACTGAGTATTGG	1980
Qy	1981	TCTCAATCATATGCAACCACTACTACCATTAACCGCTGCCACTGGTGAACACTGATACCGTT	2040
Db	1981	TCTCAATCTTATGCAACTACTACTACATCACTGCTCTCCAGGTGAACCGGATACCGTT	2040
Qy	2041	CTTATCAGAGAGCCCAACCACTACTGTCACTACTACTGAATACTGGTCTCAATCATAT	2100
Db	2041	CTTATTAGGAACCAACCAATCAGCTGTCACTACTACTGAATACTGTGTCACCAATCATTT	2100

Qy	2101	GCTAAACACCACCACCTGTTACTGCACACACCTGGTGAACCGGATACCGTTTCTTATCAGAGAG	2160
Db	2101	GCTAAACACCAACTGTAATCTGACCCACCCAGCGTGTACTGACACTGTTATCATTTAGAGAA	2160
Qy	2161	CCACAAACCATACTGCTACTACTGTAATACTGGTCTCAATCATATGCTACACACC	2220
Db	2161	CCACAAACCACACTGTCACTACTCTGAGTATTTGGTCTCAATCTTACGCAACCCACTACT	2220
Qy	2221	ACTGTTACTGCACACCAGGTGGTACCGATACTGTTATCATTTAGAGAGCCACCAAAATCCA	2280
Db	2221	ACCATTACCGCTCCACTGCTGAGACCGGATACCGTCTCTATTATAGGAAACCAACCAATCAC	2280
Qy	2281	ACAGTTACTACTGTAATATTTGGTCACAAATCATTTGGCCAAACACACACAGTTACTGCT	2340
Db	2281	ACAGTTACTACTGTAATACTGGTCACAAATCATATGCAACTACTACACCTATCATCGCA	2340
Qy	2341	CCTCAGGTGGTACTGCACCTGTGATTATCTATGA---AAGCATGTCAAAGTTCAAAGATT	2397
Db	2341	CCACAGGTGAAACTGATACTGTTTAAATCAGAGAGCCACCAACCCAACTGTTACCAACC	2400
Qy	2398	TCTACATCTTCCAATGATATACACAGTATCATTTCCATCATTTTCCGTCCTCATTTATGTC	2457
Db	2401	ACCGAATACTGGTCTCAATCCCTATACCACTGCTACTACCGTTACTTGCACCAACAGGTGA	2460
Qy	2458	AACAGCAACACCTCCGATTGTCACATTTGTAATCTTCAATCCATCAATPAC---TCCCTACT	2514
Db	2461	ACTGATACTGTGATTTATTTATGACACCATGTCAGTTTCAGAAATTTCTTCATTTTCTCGT	2520
Qy	2515	TCATACAGTAGTGATGTATGTTGTTCTTCTACAACCTTTGGTTACTGAATCAGAACA	2574
Db	2521	CCTCATTTACCAACCACTACAACCTTTGCTGTCTACAACCTGGGTATTGAAACAAAAACA	2580
Qy	2575	ACTACAGAACTGATTCGACGTGATGTAAGAGTCTCTAGATGTGCCAGTCTTCTCTGGT	2634
Db	2581	ATTACAGAACTAGCTGTGAAGGTGATAAGGTGTGTTGGTTCGTTCTTCTTCTACTCGT	2640
Qy	2635	ATTGTCACAAATCCAGATAGCAATGAATCCTCAATCGTAACTAGTATGTTTCTTACTGCA	2694
Db	2641	ATTGTCACAATTCCTAATAATATCGAAACTCCTATGGTTACTTAATACTGTTGATCTTACA	2700
Qy	2695	AGTCAATGTCGATTCATCTTCTTCAACTGATGTTATAGTGTACATCTTCTTGATAAT	2754
Db	2701	ACCAGAACTCCACTTCACAACTCCCACTC---TGGTATTTTTCAGAGTCAGGAGTATCT	2757
Qy	2755	GTTTCAAAATCAGAGTATCAGTTACAAACCGAAACTCTGTTACAACCTATTTCAAACTACT	2814
Db	2758	GTTCAAAACAGAACTTCTACTGTTACTACTGCTCAAAACAAATCCCAAGTGTCCCAACA	2817
Qy	2815	CCAAACCCATTCATCTTCAGTGACATCATTTGACTACTAGTGTGCTTTCAATTTCCAAGTGT	2874
Db	2818	GAAAGTCAGGTGTGTTATTACTACTAAGGAAACAAACGAAATGGTCTTATGAAATCACA	2877
Qy	2875	TCAGAAAGTGAAGTAAAGTTACATTTTACAACCAATGGAGACAACCAAGGTGTTACTCAT	2934
Db	2878	TCTACTAATGTGAATCAAGTATGGATGAAACTCTGAATTTA---CTTACTCCACAGCT	2934
Qy	2935	GATTACAATCTACTTCCACTGAAATGGA---ATTGTAACAAACAGGTCTTACTAAAAGTT	2991
Db	2935	GCTTCCACTTCTACTGATATTGAAATGAAACCATAGCAACAAACCGGTTCCGTCGGAAGCT	2994
Qy	2992	TTACCACTGCTGTTTCTTCTAATGACTGTTGACTAGTGAACCAACCAAAATACCAAGAA	3051
Db	2995	TCATCGGCTATCATTTCTTCTAGTGTGATGAACACTACTACTGTTTACTACTACTGCTGNA	3054
Qy	3052	CAACCAACTACATTTATCACTTCTCAAACCTCCATCACTGAAGATATCAACCACTTCTCAA	3111
Db	3055	TCAACCACTGCTATGAAACCAACCAACCAATATAATGGTGGTGAAGGCCCATCTGCCA	3114
Qy	3112	CCTACAGTGATTAATGAGACAAATACTTTCATCAACCAATCCAGTTTCCAACTGTGGCAACA	3171
Db	3115	ACTTCATCTCCATCTACAACCTACAACCTGCTAATAATGACTCTGTTATTACTGTTACTGTA	3174


```
QY 721 TGAATATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTTACATCTAATGCA 780
D 1417 TGAATATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTTACATCTAATGCT 1476
QY 781 ATTGAGATTAATATCAAAATGACTGCTGTTGTTATCGTCCATTTATGATGCTTATATT 840
D 1477 ATCTTTTACATATAAAATGTTCTGCGGTTATCGTCCATTTGTTGACGCTTATATT 1536
QY 841 TCTGCTACAGATGTTAAACCAATATCTTTAGCATATACCAATGATTTATATCTTGCTGTCG 900
D 1537 TCTGCTACAGATGTTAAACCAATATCTTTAGCATATACCAATGATTTATATCTTGCTGTCG 1596
QY 901 AGTCTGCTGCTAAAGTAACCTTTTCACTTTAAGATGGAGCTGGATACAGAATAGTGATGCT 960
D 1597 GGTATTGGCAACGCTGACCTTTTCACTTTAAGATGGAGCTGGATACAGAATAGTGATGCT 1656
QY 961 GGATCTAACGGTATTGTCATTTGTTGCTACAGTACAGTACAGTACAGTACAGTACAGTAC 1020
D 1657 GGATCTAACGGTATTGTTGCTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 1716
QY 1021 GTCACTACTTTTACCATTTCAATCCAGTGTGATATAAAACCAAAACCAATCGAAATTTTGC 1080
D 1717 GTGACCACTTACCATTTCAATCCAGTGTGATATAAAACCAAAACCAATCGAAATTTTGC 1776
QY 1081 CCTATTCCAAACCACTACCATCTGATTTGCTACAGTACAGTACAGTACAGTACAGTAC 1140
D 1777 CCTATTCCAAACCACTACCATCTGATTTGCTACAGTACAGTACAGTACAGTACAGTAC 1836
QY 1141 AGACTGTCACCAATTTGTTGATTAACAGTACTGTTATTTGTTGATGTTGCTATCATAC 1200
D 1837 AAACTGTCACCAATTTGTTGATTAACAGTACTGTTATTTGTTGATGTTGCTATCATAC 1896
QY 1201 ACAACTGTTTACAGTGAATGGAGCAACAATCACTACCAACCAACTCGTACCAATCCA 1260
D 1897 ACTACTGTTTACAGTGAATGGAGCAACAATCACTACCAACCAACTCGTACCAATCCA 1956
QY 1261 ACTGATTCATTAAGACAGTGGTGGTACAGTTCCTACTGCTACCTACTGCTACCTACT 1320
D 1957 ACTGATTCATTAAGACAGTGGTGGTACAGTTCCTACTGCTACCTACTGCTACCTACT 2016
QY 1321 ACTGATATTTGCTCAGTCCCTTGTGCTACCAACCACTACAGTTCCTCTCCAGTGGT 1380
D 2017 ACTGATATTTGCTCAGTCCCTTGTGCTACCAACCACTACAGTTCCTCTCCAGTGGT 2076
QY 1381 ACCGATCTGCTGATTTATCAGAGGACCAACCACTACTGCTACTACTGCTACTGCTACT 1440
D 2077 ACTGATCTGTTTATCAGAGACCAACCACTACTGCTACTACTGCTACTGCTACTGCT 2136
QY 1441 TCACAATCTTTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1500
D 2137 TCAGAATCTTACACTACCACTACTACTACTACTACTACTACTACTACTACTACTACT 2196
QY 1501 ATTATCAGAGAACCAACCAATCCAAGTGTCTACCAACCACTACAGTTCCTCTCAAT 1560
D 2197 ATCAATGAAGAACCTTCAATCCAAGTGTCTACCAACCACTACAGTTCCTCTCAAT 2256
QY 1561 GCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1620
D 2257 ACTACCACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 2316
QY 1621 CTCTCAACCACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1680
D 2317 CCACCAATCCAACCTGCTACTACAACAGAGTACTGCTGCTCACTACTACTACTACT 2376
QY 1681 ACTGTGACTGCTCTCCAGGAGCACTGACTCAGTAAATTTGTTGCTCCCAATCTTTAC 1740
D 2377 ACTGTGACTGCTCTCCAGGAGCACTGACTCAGTAAATTTGTTGCTCCCAATCTTTAC 2436
QY 1741 ACTGTGACTGCTCTCCAGGAGCACTGACTCAGTAAATTTGTTGCTCCCAATCTTTAC 1800
D 2437 ACTGTGACTGCTCTCCAGGAGCACTGACTCAGTAAATTTGTTGCTCCCAATCTTTAC 2496
```

```
QY 1801 CCACGAGTGGTACTGACACTGTTATCATTTAGAGAGCCACCAACCACTACTGTCTACTACT 1860
D 2497 CCACGAGTGGTACTGACACTGTTATCATTTAGAGAGCCACCAACCACTACTGTCTACTACT 2556
QY 1861 ACTGAGTATTGGTCTCAATCTGTTGCTACTACCAACCTGTTACTGGTCCACCAAGTGGC 1920
D 2557 ACTGAGTATTGGTCTCAATCTTACGCAACCACTACTACCAATTTACCGCTCCACCTGGTGAG 2616
QY 1921 ACTGATCTGTTATCATTTAGGAACCAACCAACCACTGTCACCACTACTGTAATATCTG 1980
D 2617 ACCGATCTGCTTATTTAGAGAGACCAACCAACCACTACTGTAACCACTGATGATTTGG 2676
QY 1981 TCTCAATCATATGCAACCACTACTACCAATTCACGCTCCACCTGGTGAACCTGATACCGTT 2040
D 2677 TCTCAATCATATGCAACCACTACTACCAATTCACGCTCCACCTGGTGAACCTGATACCGTT 2736
QY 2041 CTATCAGAGAGCCACCAACCACTACTGTCACCTACTGTAATGATGCTGCTCAATCATAT 2100
D 2737 CTATTTAGGAACCAACCACTACTGTCACCTACTGTAATGATGCTGCTCAATCATAT 2796
QY 2101 GTTACAACCACTACTGTTACTGCAACCACTGGTGAACCACTGTTCTTATCAGAGAG 2160
D 2797 GCTACCACTACTGTTACTGCAACCACTGGTGAACCACTGTTCTTATCAGAGAG 2856
QY 2161 CCACCAACCACTACTGTTACTGCAACCACTGGTGAACCACTGTTCTTATCAGAGAG 2220
D 2857 CCACCAACCACTACTGTTACTGCAACCACTGGTGAACCACTGTTCTTATCAGAGAG 2916
QY 2221 ACTGTTACTGCAACCACTGGTGAACCACTGGTGAACCACTGTTCTTATCAGAGAG 2267
D 2917 ACTGTTACTGCAACCACTGGTGAACCACTGGTGAACCACTGTTCTTATCAGAGAG 2963

RESULT 4
AF272027 4569 bp mRNA linear PLN 24-OCT-2000
LOCUS Candida albicans agglutinin-like protein mRNA, partial cds.
DEFINITION AF272027
ACCESSION AF272027.1 GI:10952735
VERSION AF272027.1
KEYWORDS
SOURCE Candida albicans.
ORGANISM Candida albicans.
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE
AUTHORS 1 (bases 1 to 4569)
TITLE ALS4 (agglutinin-like sequence) of Candida albicans
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4569)
AUTHORS Chen, X. and Chen, J.-Y.
TITLE Direct Submission
JOURNAL Submitted (24-MAY-2000) Shanghai Institute of Biochemistry, Yue
Yang Road 320, Shanghai 200031, China
FEATURES
source Location/Qualifiers
1..4569
/organism="Candida albicans"
/strain="SC5314"
/db_xref="taxon:5476"
1..>4569
/codon_start=1
/transl_table=12
/product="agglutinin-like protein"
/protein_id="AAG25054.1"
/db_xref="GI:10952736"
/translation="MLQFLLLSLCVSVATAKVITGIEDSFNSLTWTNAAVSYSRCPA
NPTTAVIGNSLDGATAGAGDTFTLDMPCVFKFTDQTSIDLVDGRTVATNLSAE
EFTTSSVSTCTVTMTADTKAIGTITLTPFSFSGSGSDVLANSCQFTAGINTVTE
NDGDTISATVDFKSTVASSDRILLPSLSQAVSLFPOEAGNGTSGTGMFTF
AGTGAIDCSTVHVIGISNLDNWPVISESESYTKTCTSTSLVLYONVYAGYRPFV
DAYVSVATRVSSVAMRYTNIYACVGAASVDSDFTHTMSGVSNQSGNGTIVVTRTV
TDSSTAVTLPENSESDTKTIEILOPIPTTTTTSYGVGVTTSYTKTAPIGETATVI
VDVPYHTTTTTSWTGTTTITTTTNTPTDSDITVVQVPSNPPTVTTEYWSQSYAT
TTVTAPPGTDSVIRFPNPTVTTTTEYWSQSFATTTTATPGGTDSVIRFPNP
```


NDGSKLSIAVNFESTVDRSGLTTSRPMPLNKIATLYVAPOCENGYSCTGCMFST
SYGDVAIDCSNVHIGISKGVDNHNHPTSESEYTKSCSEFISITYONVPAGYRPF1
DAYLSPDSNVOLYSKYKNDYTCVDYDWOHAPFTLKWGYNKSDAGSNGIIVATRTV
TDSITAVTTLFPNSVDKTKIEILQPIPTITITISYGVTSISKTAPGETALI
VDVPVHTTTSVTLVSKWGTITMTTTRNPTDSIDTVVQVPLNPPTTTTQWSEFIS
TTTTTHKPEGTSTVIVKPHNPTVTTTFWSEYATTEITITPLGTDSDIHDPLEE
SSSTAIEDSDNSISSAQESSVSQSSIVGLSSSDIPLSMPSSSSTGLTLEE
SSTVSDSDSSSELSTFSSSESISIDTTFWSDSSSDLESTISITWSSSIDAQ
SSOSVOSNISLSTSOETTSAGEENSTVTDILVSDASSILNSDISSVPSSTISL
SDDPHTLAGPDRSSSIASSTVEISDLVSLTSDPTSSRDSSSSLSDSSSPSD
ESDLSASSFSLVAPSFSLSSSSLSLIYHYVNSTTYHASESSSVASPSVASES
ANDPHTLSESTDSITIGTDSSTVFCRRONGDCIVITGTSIDSEQSDVTTTS
SFVASSSTAEQSTIDNPNDSSQTSASSTKSSVSVDVWVNSILLESSTLSSDD
GTSDTSITLSTDPGNTAGNSHKSTKESIOKTVTLSSSYLSLSTLSSDI
TIELTTELITIEDNEPNTFTSPSSHSEIFSSDNLVSKOVDRSTLKTSPYTDVTT
VSSLSVHSEASTATLGENSPSNVASTPLNTATSLRSTSSSNHATESGIVKSEASV
EAPSPFTSSTSTEEAREGITYANGSTNNLITESQVAAPDSQVPSSSHSELST
STFDNSAAVDQPSKSTIEESIMNPDSNTETNGFIATLSQAQVPSSSHSELST
TTAKTTDASMGDSASNSQPTTLIQVATSYNOPLITTVAGSSSATKHPFWLLKFI
SVALPFFL

BASE COUNT 1172 a 819 c 641 g 1181 t

ORIGIN

Query Match 31.4%; Score 1188.8; DB 8; Length 3813;
Best Local Similarity 78.8%; Pred. No. 7.9e-205;
Matches 1418; Conservative 0; Mismatches 382; Indels 0; Gaps 0;

QY 1 ATGCTCAACAATTTACATGTTATTCCTATATTGCTCAATTCGAAGTGCAGACAATC 60
DB 1 ATGATTCAACAATTTACATGTTATTCCTATATTGCTGTTGCGACTGCAAGGCGATC 60
QY 61 ACTGGTGTGTTTGAATGTTTAACTTAACTTGTGCTCAAGTCTGCTAATATGCTTTC 120
DB 61 ACTGGTGTGTTTGAATGTTTAACTTAACTTGTGCTCAAGTCTGCTAATATGCTTTC 120
QY 121 AAAGGCCAGGATACCAACTTGGAAATGCTGTTTGGGTTGGTGCCTTAGATGGTACCAGT 180
DB 121 AAAGGACCAGGATACCAACTTGGAAATGCTGTTGGTGGTGCCTTAGATGGTACCAGT 180
QY 181 GCAATCCAGGGATACATTCATTAATATGCCATGCTGCTGTTTAAATATACACTTCA 240
DB 181 GCAATCCAGGGATACATTCATTAATATGCCATGCTGCTGTTTAAATATACACTTCC 240
QY 241 CAACAATCTGTTGATTAACTGCCGATGGTGTAAATATGCTACTTGTCAATTTTATCT 300
DB 241 CAAAAATCTGTTGATTGACTGCCGATGGTGTAAATATGCTACTTGTCAATTTTATCT 300
QY 301 GGTGAAGAAATTCACAATTTTCTACATTAACATGATGCTGAACGACGCTTTGAAATCA 360
DB 301 GGTGAAGAGTTTACAACATTTTCTACATTAACATGATGCTGAACGACGCTTTGAAATCA 360
QY 361 TCATTAAGGCATTTGGTACAGTTACTTTACCAATTTGCAATTCATGCTGGTGAACAGGT 420
DB 361 TCATTAAGGCATTTGGTACAGTTACTTTACCAATTTGCAATTCATGCTGGTGAACAGGT 420
QY 421 TCATCAACTGATTGGAAGATTCTAAATGTTTACTGCTGGTACCATACAGTCACATTT 480
DB 421 TCATCAACTGATTGGAAGATTCTAAATGTTTACTGCTGGTACCATACAGTCACATTT 480
QY 481 AATGATGGTGAATAAGATATCTCAATGATGTTGAGTTGTTGAAAGTCAACCGTTGATCCA 540
DB 481 AATGATGGGAGTAAAAAGCTCTCAATGCTGTTAAATTTGAAAGTCAACAGTTGATCCA 540
QY 541 ACTGCATATTTGATGCTTCCAGAGTTATGCCAAGTCTCAATAAGGTCACAACCTCTTTT 600
DB 541 AGTGGGTATTTGACTACTTCCAGATTATGCGGAGTCTCAATAAAATTTGCTACTCTTAT 600
QY 601 GTGGCACCACAATGTGAAAATGGTTACACATCTGGTACAATGGGTTCTCCAGTAGTAAC 660
DB 601 GTGGCACCACAATGTGAAAAGGTTACACATCTGGTACAATGGGTTCTCCAGTAGTAT 660
QY 661 GTGACGTTGCTATGATTGCTCAAAATATTCATATTTGGTATATCAAAAAGGATTAATGAT 720
DB 661 GTGACGTTGCTATGATTGCTCAAAATATTCATATTTGGTATATCAAAAAGGATTAATGAT 720

DB 661 GGGGATGTTGCTATTGACTGTTTCAAATGTACATATTTGGTATTTCAAAAGGAGTAATGAT 720
QY 721 TGGAAATATCCGGTTCATCTGAATCATTTAGTTTACATTAACATTTGTACATCTAATGGA 780
DB 721 TGGAAATATCCGGTTCATCTGAATCATTTAGTTTACATTAACATTTGTACATCTAATGGA 780
QY 781 ATTCAAGTAAATATCAAAATGACTGCTGCTGTTTATCGTCCATTTATTTGATGCTTATAT 840
DB 781 ATCTCTATCATATCAAAATGTTCTCGCGGTATTCGTCATTTATTTGACGCTTATAT 840
QY 841 TCTGTACAGATGTTAAACCAATATCTTTAGCATATACCAATGATTTACTTGTGCTGCG 900
DB 841 TCTCCCTCATGAATAAACAGTATCAATTCGTCGTAATAAATGACTATACATCTTGTGTTGAT 900
QY 901 AGTCGCTCTCAAAAGTAAACCTTTCACTTTAAGATGACTGCTGATACAGATACGATGATGCC 960
DB 901 GATTATTGGCAACATGACCTTTCACTTTAATAATGGACTGGATATAGAATAGTATGATGCC 960
QY 961 GGATCTAAAGGTTATTTGCTATTTGCTTACAACCTAGAACAGTTACAGACAGTACCACTGCT 1020
DB 961 GGATCTAAAGGTTATTTGCTATTTGCTTACAACCTAGAACAGTTACAGACAGTACCACTGCT 1020
QY 1021 GTCACACTTTTACCATTCAATCCAAAGTGTGATATAAACCACCAATCGAAATTTTGCAC 1080
DB 1021 GTCACACTTTTACCATTCAATCCAAAGTGTGATATAAACCACCAATCGAAATTTTGCAC 1080
QY 1081 CCTATTCCAAACCACTACCACCACTTCATATGTTGGTGGTGAATCTCTCTCTATCTGACT 1140
DB 1081 CCTATTCCAAACCACTACCACCACTTCATATGTTGGTGGTGAATCTCTCTCTATCTGACT 1140
QY 1141 AAGACTGCAACCAATTTGGTGAACAGCTACTGTTTATTTGTTGATGCTGCAATATCATCTACC 1200
DB 1141 AAGACTGCAACCAATTTGGTGAACAGCTACTGTTTATTTGTTGATGCTGCAATATCATCTACC 1200
QY 1201 ACAACTGTTTACCAGTGAATGGAGCAGAACAACTACTACCAACCACTCTCTACCAATCCA 1260
DB 1201 ACAACTGTTTACCAGTGAATGGAGCAGAACAACTACTACCAACCACTCTCTACCAATCCA 1260
QY 1261 ACTGATTCATTCACACAGTGGTGGTACAGTTCCCACTGCGCAATCCAACTGTTTAGTACT 1320
DB 1261 ACTGATTCATTCACACAGTGGTGGTACAGTTCCCACTGCGCAATCCAACTGTTTAGTACT 1320
QY 1321 ACTGAATATTTGCTCAGTCTCAGTCTTGTGTACAAACCACTACTGCTCTCTCTCCAGTGGT 1380
DB 1321 ACCCAGTTTTTGGTGCAGAGTCATTTTACTAGTACTACTACCAATCACAACCAAGCAGAAGC 1380
QY 1381 ACCGATCTGTGATTTATCAGAGGACCAACCACTACTGTCTACTACTACTACTGATATTGG 1440
DB 1381 ACAGACTCAGTCAATGTCAGAGAACCAACCACTACTGTCTACTCTCTCCACACAGAGTTTGG 1440
QY 1441 TCACAATCTTCTGCTACTACTACTGTTTACTGCTCTCCAGTGGTACTGACTCAGTA 1500
DB 1441 TCAGATCATATGCCACTACTGTAACANATTTACTACAGGCGCACTTGGTACTGATGATATC 1500
QY 1501 ATTATCAGAGAACCAACCAATCCAACTGCTACTACCAACCAAGATTTGGTCTCAATCTCTT 1560
DB 1501 GTTATACATGATCCATTTGGAGAGAACGTCTCTCTACTACTGCTATTTAGTGCAGTGATCT 1560
QY 1561 GCTACTACTACTACAGTTACTGCTCTCCAGTGGTGGTACTGACTCAGTAATTTATCAGAGAA 1620
DB 1561 AATATTTCAAGCTCAGCTCAAGAAATCATCCAGCTGCTGGTGGAGGATCATCAAGTATATG 1620
QY 1621 CCTCCAAACCACTGTCACCACTGTAATATTTGGTCCCAATCTTACGCAACCAACAT 1680
DB 1621 GGATTTGATCAAGTTTCAAGATATACCAATTAAGTTTCAAGATGCTCATCGAGTCAACT 1680
QY 1681 ACTGTGACTGCTCTCTCAGGAGGCACTGACTCAGTAATTTATCAGAGAACCAACCAAC 1740
DB 1681 GGTTTAACATCTAGTGTGAGTCTCTACTGCTCAAGTTATGATAGCGATTCATCAAGTAGT 1740
QY 1741 ACTGTCACTACTACTGTAATGATGCTGTCACAATCATATGCCACCACTACTGATCACTGCA 1800
DB 1741 AGTGAGTTATCTACATTTTCTAGTTCTGAGAGGCTACTGCTCAAGTATCTCTGATACACA 1800

RESULT 7
AR044076
LOCUS AR044076 1071 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 1 from patent US 5817466.
ACCESSION AR044076
VERSION *AR044076.1 GI:5965541
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1071)
AUTHORS Hoyer, L.L., Livi, G.P. and Shatzman, A.R.
TITLE Conserved yeast nucleic acid sequences
JOURNAL Patent: US 5817466-A 1 06-OCT-1998;
FEATURES Location/Qualifiers
source 1..1071
BASE COUNT 317 a 322 c 159 g 273 t
ORIGIN
Query Match 28.2%; Score 1069.4; DB 6; Length 1071;
Best Local Similarity 99.9%; Pred. No. 3.8e-183;
Matches 1070; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1300 CCAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCTTGTGTACAAACCACTACA 1359
Db 1 CCAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCTTGTGTACAAACCACTACA 60
Qy 1360 GTTACTGCTCTCCAGGTGGTACCGATACGTGATTATCAGAGAGCCACCAACCACTACT 1419
Db 61 GTTACTGCTCTCCAGGTGGTACCGATACGTGATTATCAGAGAGCCACCAACCACTACT 120
Qy 1420 GTCACTACTACTGAATATTGGTCAACATCCTTCTCTACTACTACTGTTACTGCTCCT 1479
Db 121 GTCACTACTACTGAATATTGGTCAACATCCTTCTCTACTACTACTGTTACTGCTCCT 180
Qy 1480 CCAGGTGGTACTGACTCAGTAATTTATCAGAGAACCCAAATCCAACTGCTACTACAAC 1539
Db 181 CCAGGTGGTACTGACTCAGTAATTTATCAGAGAACCCAAATCCAACTGCTACTACAAC 240
Qy 1540 GAGTATTGGTCTCAATCCTTTGCTACTACTACTACTGCTCTCCAGGTGGTACT 1599
Db 241 GAGTATTGGTCTCAATCCTTTGCTACTACTACTACTGCTCTCCAGGTGGTACT 300
Qy 1600 GACTCAGTAATTTATCAGAGAACCTCCAAACCACTGTCACCACTGTAATATTGGTCC 1659
Db 301 GACTCAGTAATTTATCAGAGAACCTCCAAACCACTGTCACCACTGTAATATTGGTCC 360
Qy 1660 CAATCTTACGCAACCACTACTGCTGACTGCTCTCCAGAGGCACTGACTCAGTAATT 1719
Db 361 CAATCTTACGCAACCACTACTGCTGACTGCTCTCCAGAGGCACTGACTCAGTAATT 420
Qy 1720 ATCAGAGAACCCAAACCACTGCTACTACTACTGAATATTGGTCAACATCATATGCC 1779
Db 421 ATCAGAGAACCCAAACCACTGCTACTACTGTAATATTGGTCAACATCATATGCC 480
Qy 1780 ACCACTACCACTGTAACCTGCACCACTGCTGACTGCTGTAATATTGGTCAACATCATATGCC 1839
Db 481 ACCACTACCACTGTAACCTGCACCACTGCTGACTGCTGTAATATTGGTCAACATCATATGCC 540
Qy 1840 CCAACCACTGCTACTACTACTGAGTATTGGTCTCAATCGTTTGGTACTTACCAACT 1899
Db 541 CCAACCACTGCTACTACTACTGAGTATTGGTCTCAATCGTTTGGTACTTACCAACT 600
Qy 1900 GTAAGTGTCCACCAAGTGGCACTGATCTGTTATCATATTAGGGAACCCAAACCACT 1959
Db 601 GTAAGTGTCCACCAAGTGGCACTGATCTGTTATCATATTAGGGAACCCAAACCACT 660
Qy 1960 GTCACCACTACTGTAATCTGCTCAATCATATGCAACCACTACTAGCAATTTACCGTCCA 2019
Db 661 GTCACCACTACTGTAATCTGCTCAATCATATGCAACCACTACTAGCAATTTACCGTCCA 720

Qy 2020 CCTGGTGAAGTATACCGTTCTTTATCAGAGAGCCACCAACCACTACTGTGCTACTACT 2079
Db 721 CCTGGTGAAGTATACCGTTCTTCATCAGAGAGCCACCAACCACTACTGTGCTACTACT 780
Qy 2080 GAATACCTGGTCTCAATCATATGCTTACACCACTACTGTTACTGCAACCACTGGTGAAC 2139
Db 781 GAATACCTGGTCTCAATCATATGCTTACACCACTACTGTTACTGCAACCACTGGTGAAC 840
Qy 2140 GATACCGTCTTATCAGAGAGCCACCAACCACTACTGTGCTACTACTGAATATTGGTCT 2199
Db 841 GATACCGTCTTATCAGAGAGCCACCAACCACTACTGTGCTACTACTGAATATTGGTCT 900
Qy 2200 CAATCATATGCTACAAACCACTGTTTACTGCAACCACTGGTGGTACTGTTTATC 2259
Db 901 CAATCATATGCTACAAACCACTGTTTACTGCAACCACTGGTGGTACTGTTTATC 960
Qy 2260 ATTAGAGAGCCACCAATCCAACTACTACTACTGTAATATTGGTCAACATCATTTGCC 2319
Db 961 ATTAGAGAGCCACCAATCCAACTACTACTACTGTAATATTGGTCAACATCATTTGCC 1020
Qy 2320 AACACCACTGTTACTGCTCTCCAGGTGGTACTGACACTGTGATTATC 2370
Db 1021 AACACCACTGTTACTGCTCTCCAGGTGGTACTGACACTGTGATTATC 1071
RESULT 8
165632
LOCUS 165632 1071 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 1 from patent US 568263.
ACCESSION 165632
VERSION 165632.1 GI:2482202
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1071)
AUTHORS Hoyer, L.L., Livi, G.P. and Shatzman, A.R.
TITLE Conserved yeast nucleic acid sequences
JOURNAL Patent: US 568263-A 1 16-SEP-1997;
FEATURES Location/Qualifiers
source 1..1071
BASE COUNT 317 a 322 c 159 g 273 t
ORIGIN
Query Match 28.2%; Score 1069.4; DB 6; Length 1071;
Best Local Similarity 99.9%; Pred. No. 3.8e-183;
Matches 1070; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1300 CCAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCTTGTGTACAAACCACTACA 1359
Db 1 CCAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCTTGTGTACAAACCACTACA 60
Qy 1360 GTTACTGCTCTCCAGGTGGTACCGATACGTGATTATCAGAGAGCCACCAACCACTACT 1419
Db 61 GTTACTGCTCTCCAGGTGGTACCGATACGTGATTATCAGAGAGCCACCAACCACTACT 120
Qy 1420 GTCACTACTACTGAATATTGGTCAACATCCTTCTCTACTACTACTGTTACTGCTCCT 1479
Db 121 GTCACTACTACTGAATATTGGTCAACATCCTTCTCTACTACTACTGTTACTGCTCCT 180
Qy 1480 CCAGGTGGTACTGACTCAGTAATTTATCAGAGAACCCAAATCCAACTGCTACTACAAC 1539
Db 181 CCAGGTGGTACTGACTCAGTAATTTATCAGAGAACCCAAATCCAACTGCTACTACAAC 240
Qy 1540 GAGTATTGGTCTCAATCCTTTGCTACTACTACTACTGCTCTCCAGGTGGTACT 1599
Db 241 GAGTATTGGTCTCAATCCTTTGCTACTACTACTACTGCTCTCCAGGTGGTACT 300
Qy 1600 GACTCAGTAATTTATCAGAGAACCTCCAAACCACTGTCACCACTGTAATATTGGTCC 1659
Db 301 GACTCAGTAATTTATCAGAGAACCTCCAAACCACTGTCACCACTGTAATATTGGTCC 360

JOURNAL Submitted (11-SEP-1997) Veterinary Pathobiology, University of Illinois, 2001 S. Lincoln Avenue, Urbana, IL 61802, USA

FEATURES

Location/Qualifiers

1. .1404

/organism="Candida albicans"

/strain="1161"

/db_xref="taxon:5476"

/map="6C"

gene 1. .>1404

/gene="ALS2"

/note="agglutinin-like sequence"

/allele="ALS2-2"

CDS 1. .>1404

/gene="ALS2"

/codon_start=1

/transl_table=12

/product="agglutinin-like protein"

/protein_id="AAC64237.1"

/db_xref="GI:3598675"

/translation="MLQLFLLSLCVSVATAKVITGVFNSFDSLTTWTRAGNYAKGNP
RPTNAVLGWSLDGTSANPDGFTLLNMPGVFKFIDQTSVDLTAEGVKATCFYSGE
EDFTSSLKTSVNTLTSSIKALGTVTLPSFNVGGTGSVDLESQCRKAGNTVTF
NDGDKIISIDVDFEKTNDISAGFFIASRLIPSLINKVSIYVAPQCANGYTSGAMGT
LVGDTTIDCSNVHVGITKGLDNFVPSDDLSLYNKTCSTGIGISYENVPAGRYRFF
DYVTSVGNQRLPNRYDVCVSSLOSXPFLRGLKYNSEANSNGFVLIVATRTVIV
DSTTAVTLQFNSVDKTKTIEILQIPITTTITSYGVGTTSYSTKTAIGETATVIV
DVPYTGTTVTSWMTGTITATTTTNTPTDSIDTVVVQVPSNPVTSTTEYSQSYATV
TWTVPAGGTGDSVLIIEP"

repeat_region 1297 -1404

/note="first of multiple copies of 108bp imperfect repeat"

BASE COUNT 421 a 266 c 255 g 462 t

ORIGIN

Query Match	26.9%;	Score 1016.6;	DB 8;	Length 1404;
Best Local Similarity	83.2%;	Pred. No. 1.2e-173;		
Matches 1170;	Conservative	0;	Mismatches 234;	Indels 3; Gaps 1;

```

QY      1  ATGCTTCAACAATTACATTTGTTATTCCTATATTGTTCAAATGCGAAGTGCACAAACAAATC 60
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D6      1  ATGCTTTTACAATTTTGTGTGCTAAGCCTCTGCTATCAGTTGCTACTGCGAAAGTTATT 60
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY      61  ACHGTGTTTTTGGATAGTTTTTAATTCATTAACCTGGTGCCAAATGCTGCTAATATGCTTTC 120
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D6      61  ACGGGTGTTTTCAATAGTTTTGTATTCGTTGACATGGACAAGAGCTGGTAAATATGCTTAT 120
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY      121  AAAGGCCAGGATACCAACTTGGAAAGCTGTTTGGTTGGCTCTAGATGGTACCAAGT 180
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D6      121  AAGGGCCCAANTAGCCAACCTTGAATGCTGTTTTGGGTGGCTCTTAGATGGTACTAGT 180
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY      181  GCCAATCCAGGGGATACATTCACATTCGAATATGCCATCTGTGTTTAAATATACTACTTCA 240
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D6      181  GCAATCCAGGAGACACATTCACATTCGAATATGCCATGTGTTTTAAATTTATACCGAT 240
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY      241  CAACATCTGTTGATTTAACTGGGATGGTGGTAAATATGCTACTGTGCAATTTTATTTCT 300
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D6      241  CAAACATCTGTTGATTTGACTGCTGAAGGTGTTAAATATGCTACATGCTAGCTTTTATTC 300
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY      301  GGTGAAGAAATTCACAACTTTTCTACATTAACATGTACTGTGAACGAGCTTTGAAATCA 360
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D6      301  GGTGAAGAAATTCACAACTTTTCTTCATTAATAATGTACTGTGAGCAATACATTTTAACTCA 360
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY      361  TCCTATTAGGCATTTGGTACAGTTTACTTTACCAATTCGATTTGTTGGTGGAAACAGGT 420
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D6      361  TCATTTAAGCTTTGGGTACGGTTACTTTACCAATTCATTTAATGTTGGTGGAAACAGGT 420
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY      421  TCATCAACTGATTTGGGAAGATTTCTAAATGTTTTTACTGCTGTTACCAATACAGTACACATT 480
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D6      421  TCATCGGTTGATTTGGGAAGTTCTCAATGTTTTTAAGGCTGCGACCAACACAGTTACTTTT 480
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY      481  AATGATGGTGATAAAGATATCTCAATTTGATGTGAGTTTGAAGAAGTCAACCGTTGATCCA 540
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```
Db 481 AATGATGGTGATAAAAAATTTCAATTGACGCTTGATTTTGAAAAAACAACGAAGATGCA 540
Qy 541 AGTCATATTTGTTGTTGCTTCCAGAGTTATGCAAGTCTCAATAAGGTCACAACTCTTTT 600
Db 541 AGTGAATTTTATGCGTCAAGACTTATCCAGATTAACAAGTTTCAATCACTTAT 600
Qy 601 GTGCACACAAATGTGAATGGTTACACATCTGGTCAATGCGGTCTCCAGTAGTAAC 660
Db 601 GTGCACACAAATGTGAATGGTTCACATCTGGTCAATGCGGTCTCATAGTTCTCACT 660
Qy 661 GGTGACGTTGCTATGTTGCTCAAAATTCATATTTGGTATCACAAAGGATTAATGAT 720
Db 661 GGTGACACTATGTTGCTGTTCAAAATGTTCAATGTTGGTATCACAAAGGATTAATGAT 720
Qy 721 TGAATATTCGGGTTTCATCTGAATCATTTAGTTTACACTAAACTCTGTACATCAATGGA 780
Db 721 TGGATTTTCGGGTTTCGTCGATTCATTAAGTTTACAAATAAACTTGTTCATCTACAGGT 780
Qy 781 ATTGAGATTAATATCAAAATGTACCTGCTGGTTATCGTCCATTTATTTGATGCTTAT 840
Db 781 ATTTCTATCATATGAAATGTCCCGCTGGTTATCGTCCATTTTGTGACGTATATACT 840
Qy 841 TCTGCTACAGATGTTAACCAATATACCTTACATATACCAATGATTTACTTGTCTGGC 900
Db 841 CTGGTGTGACGCCGAGAACAGACA---ATTAGATATATACTAATGATTTGCGTGTGGT 897
Qy 901 AGTCCTCTGCAAAAGTAAACCTTTCACTTTTAAGATGGACTGGATACAGAAATAGTATGCC 960
Db 901 AGTCCTCTGCAAAAGTAAACCTTTCACTTTTAAGATGGACTGGATACAGAAATAGTATGCC 960
Qy 957 AGTTCCCTTACAAAGTAAAGCGGTTCAATTTAAGATTGAGAGGATACAAATAGTGAAGCT 957
Qy 1021 GTCACTACTTTTACCATTCAATCCAAAGTGTGATAAAACCAAAACAATCGAAATTTTGCAA 1080
Db 1018 GTCACTACTTTTACCTTTTATCCAAAGTGTGACAAACCAAAACAATCGAAATTTTGCAA 1077
Qy 1081 CCTATTCACCACTACATACACAACTTCATATGTTGGTGTGACTTCTCTATCTGACT 1140
Db 1078 CCTATTCACCACTACATACACAACTTCATATGTTGGTGTGACTTCTCTATCTGACT 1137
Qy 1141 AAGACTGACCAATTTGGTGAACAGCTACTCTTATTTGATGTGCTATATCATACTACC 1200
Db 1138 AAGACTGACCAATTTGGTGAACAGCTACTCTTATTTGATGTGCTATATCATACTACC 1197
Qy 1201 ACAACTGTTACAGTGAATGGACAGGAACAATCACTACCAACCAACTCGTACCAATCCA 1260
Db 1198 ACAACTGTTACAGTGAATGGACAGGAACAATCACTACCAACCAACTCGTACCAATCCA 1257
Qy 1261 ACTGATTCATATGACAGTGGTGGTACAAGTTCCTACTGCCAAATCCAACTGTTAGTACT 1320
Db 1258 ACTGATTCATATGATAGTACTGCTGTTGCAAGTTCCTACTGCCAAATCCAACTGTTAGTACT 1317
Qy 1321 ACTGAATATTTGGTCTCAGTCTTGTCTACACCACTACAGTTACTGCTCTCCAGGTGGT 1380
Db 1318 ACTGAATATTTGGTCTCAGTCTATGCTACACCACTACAGTTACTGCTCTCCAGGTGGT 1377
Qy 1381 ACCGATCTGTGATTTATCAGAGGCCA 1407
Db 1378 ACTGACTCAGTCATATACAGAGGCCA 1404

RESULT 10
CALALS1
LOCUS
DEFINITION
  Candida albicans agglutinin-like protein (ALS2) gene, 5' partial cds.
ACCESSION
  AF024580
VERSION
  AF024580.1
KEYWORDS
  GI:3598667
SEGMENT
  1 of 2
SOURCE
  Candida albicans.
```

```
ORGANISM
  Candida albicans
  Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
  Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE
  1 (bases 1 to 1404)
AUTHORS
  Hoyer, L.L., Payne, T.L. and Hecht, J.E.
TITLE
  Identification of Candida albicans ALS2 and ALS4 and localization
  of als proteins to the fungal cell surface
JOURNAL
  J. Bacteriol. 180 (20), 5334-5343 (1998)
MEDLINE
  98440424
PUBMED
  9765564
REFERENCE
  2 (bases 1 to 1404)
AUTHORS
  Hoyer, L.L.
TITLE
  Direct Submission
JOURNAL
  Submitted (11-SEP-1997) Veterinary Pathobiology, University of
  Illinois, 2001 S. Lincoln Avenue, Urbana, IL 61802, USA
FEATURES
  Location/Qualifiers
    1..1404
      /organism="Candida albicans"
      /strain="1161"
      /db_xref="taxon:5476"
      /chromosome="6"
      /map="6; C"
      /gene="ALS2"
      /codon_start=1
      /transl_table=12
      /product="agglutinin-like protein"
      /protein_id="AAC64235.1"
      /db_xref="GI:3598670"
      /translation="MLQFLLSLCVSATAKVITGVNFSFDSLTVTRAGNYAKGN
      RPTWNAVLGMSLDGTSANPGDITLNNPCVFKFITDQTSVDLTAEGVKATCFYSGE
      EFTFSSLKTSVNTLTSTKALGTVLPISFNVGGTSSVDLESQCFKAGTNTVTF
      NQDGKIKSIDVDEKTNEDASGYFIASRLIPSIKVSITYVAPQCANGYTSAGMGFTV
      LCTGDTTDCSNVHGITTGLNDWNFPVSSDSLSYKNTCSSTGTSITVENYVAGYRPF
      DVTYSVGNQRLRYNDYACVGSLSQSKPFLNRLRGYNNSEANSNGFVIVATRTVT
      DSTAVTLPFPNESVDKTKTIEILOPITTTITSYGVGTTSTKTAPIGERTATV
      DVPHTTTTISEWTGTTTTTTTTRNPTDSDIVVVOVPSFNTVTITTEWSQSIAT
      TTVTAPPGGTDVSIIREP"
    repeat_region
      1297..1404
      /note="first of multiple copies of imperfect 108 bp
      repeat"
    BASE COUNT
      422 a 268 c 253 g 461 t
    ORIGIN
      Query Match
      Best Local Similarity 26.6%; Score 1006; DB 8; Length 1404;
      Matches 1163; Conservative 0; Mismatches 240; Indels 3; Gaps 1;
      Qy 1 ATGCTTCAACAATTTACATGTTTATCTCTATATTTGCAATTTGCAAGTGCACAAACAATC 60
      Db 1 ATGCTTTTACAATTTTGTGCTAAGCCTCTGTGTATCAGTTGCTACTGCAAAAAGTTATT 60
      Qy 61 ACTGCTGTTTGTAGTGTATTAATTCATTAATTCATTTGGTCCCAATGCTGCTTAATATGCTTTC 120
      Db 61 ACGGCTGTTTTCATAGTTTGTGATTTGATTTGATGTCGATGGAAGAGCTGTTAATATGCTTAT 120
      Qy 121 AAAGGGCCAGGATACCCAACTTGGAAATGCTGTTTGGTGTGCTCTAGATGTTGACAGT 180
      Db 121 AAGGCCCAATAGACCAACTTGGAAATGCTGTTTGGCTGCTCTTTAGATGTTAGT 180
      Qy 181 GCCAATCCAGGGATACATTCACATTTGAATGCAATGTCATGTTGTTTAAATATACACTTCA 240
      Db 181 GCAATCCAGGAGACACATTCACATTTGAATGCAATGTCATGTTTAAATTTATTACCGAT 240
      Qy 241 CAACATCTGTTGATTTAACTGCCGATGGTGTAAATATGCTACTTGTCAATTTTATTTCT 300
      Db 241 CAACATCTGTTGATTTGAGCTGCGAAGGTTGTTAAATATGCTACTGTCAGTTTATTTCA 300
      Qy 301 GGTGAAGAATTCACAACCTTTTCTTACATTAACATGCTACTGTGAACGCGCTTTGAAATCA 360
      Db 301 GGTGAAGAATTTACAACATTTTCTTCAATAAATGCTACTGTGACCACTACTTTTACATCA 360
```


Db 26 TTTTCTCTATTGACGATAGCAATGGCAAAACATATATCGGGAGTTTTCACGAGTTTCA 85
Qy 83 ATTCAATAACTTGGTCCAAATGCTCTAATATATGCTTTCAAGGGCCAGGATACCAACTT 142
Db 86 ACTCATGGACCTACTAATATGCTGTAATACCATATGGGGGCTCGGTATATCCACAT 145
Qy 143 GGAATGCTGTTTGGGTGGTCCCTTAGATGGTACCGAGTCCCAATCCAGGGGATACATTCA 202
Db 146 GGACTGCTGTTTGGGTGGAGCTTGGAGGGAACACACTAGCTAGTCCAGGTGATACATTTA 205
Qy 203 CATTGAATATGCCAATGCTGTTTAAATATATACTACTTTCACAAACATCTGTTGATTTACTG 262
Db 206 CATTGGTCAATGCCCTCGGTTTCAAAATTTATACCAACAACTTCAGTACGATTAAGT 265
Qy 263 CCGATGGTGTAAATATGCTACTTGTCAATTTTATCTGGTGAAGAATTCACAACCTTTT 322
Db 266 CTATGGTGTCAAGTATGCAACATGTAATTTCCATGAGGGGAACACTTTACTACTTTT 325
Qy 323 CTACATTAACATGTAATGTAACGACGCTTTGAAATCATCCATTAAGGCAATTTGGTACAG 382
Db 326 CAAGTATGAGTGTGTAGTAATAATGGGCTATCTTCAAAATATCAGAGCGTTTGGTACCG 385
Qy 383 TTACTTTACCAATTTGCATTCATGTTGGTGAACAGGTTTCATCAACTGATTTGGAAGATT 442
Db 386 TCAGGCTACCAATTTGCATTCATGTTGGTGAAGTGTTCATCTGTCACCAATTCAGATT 445
Qy 443 CTAAATGTTTACTGCTGGTACCAATACAGTCACATTTAATGATGGTGAAGAATATCT 502
Db 446 CAAAGTGTTCACGTGCGAAGCAACTGTAAACATTTACAGACGCGGATCACAAATTT 505
Qy 503 CAATGTATGTAAGTTGAAAGTCAACCGTTGATGTAACGAGTATTTGATGCTTCCA 562
Db 506 CTACTACAGTCAATTTCCCTAAGACTCCACAATCATCTAGTAGTGTGGTTTATTCGCA 565
Qy 563 GAGTTATGCCAAGTCTCAATAAGGTCAACAATCTTTTGTGGCCACCAATGTGAAATG 622
Db 566 GGGTTTTCCTCAAGTCTTGATAAATATCTAGTCTTGTGTTCTCAGGTACTGCTG 625
Qy 623 GTTACATCTGTTACATAGGGGTTCTCCAGTAGTACGGTACGTTGCTATTTGATTTGCT 682
Db 626 GATATGATCCGGTGTGCTCGGATTTTCAGCAACAAAGATGATGTGACAAATTCATTGT 685
Qy 683 CAAATATTCATATTTGGTATCACAAAGATTAATGATGTTGGAATTTATCCGGTTTCATCTG 742
Db 686 CTACTATACATGTTGGGAATAACAAATGGTTGGAATAGTTGGAATATCCAGTATCATCAG 745
Qy 743 AATCATTTAGTTAGTACAAATTTGTACATCTAATGGAATTCAGATTAATATCAAAATG 802
Db 746 AATCATTTTCTACACCAAACTTTGACACCAACAGTTTATTTATTTACTTTATGAAATG 805
Qy 803 TACCTGCTGGTTATCGTCCATTTATGATGCTT-----ATATTTCTGCTACAGATGTTA 856
Db 806 TTCCTGCAAGTTATCGTCCATTTATGATTTCTACGTGAAAAAATCAGCAACGAACGA 865
Qy 857 ACCAATATCTTTAGCATATACCAATGATTTATCTTGTGCTGGCAGTCTGCTGCAAGTA 916
Db 866 ATGATTTAATTTGAATTTACACGAATATATCAAAATTTGATGGATGGCAAAAAGGGAATG 925
Qy 917 AACCTTTTCACTTTAAGATGGACTGGATCAAGAATAGTGTCCGGGATCTAACGGTATTG 976
Db 926 ATCCTCTTATATCTTTGGACATCATACAAATAGTATGATGAGGATCCCAATGGAGCTG 985
Qy 977 TCATTTGTTGCTACAACTAGAACAGTTACAGACAGTACCACGTCGTCTCACTACTTTACCAT 1036
Db 986 CCGTAGTTGTTTACTACGAGAAGTCACTGATTTCTACAACAGCAATTAACCAATTAACCGT 1045
Qy 1037 TCATCCCAAGTTTGAATAAACCAAAACATCAAGAAATTTTCCACCTATTTCCACCACTA 1096
Db 1046 TTGATCCCAAGTTTGAATAAACCAAAACATTAAGTAATAGAACCCATCCCTACTACCA 1105
Qy 1097 CCATCAACAACTTCATATGTTGGTGTGACTTCTCTCTGACTAAGACTGACCAAAATG 1156
Db 1106 CTATTACTACTCATATGTTGGGATTTCTACTTCTTCTACGAAAGACTGCAACTATG 1165

Qy 1157 GTGAAACAGCTACTGTTATTTGATGTGCAATATCATATACCACAACTGTTTACCAGTG 1216
Db 1166 GAGAACAGCAACTGTTGTTGTTGATGTTCCTATATACAACTACCACCTATCATAGTA 1225
Qy 1217 AATGACAGGAACAATCACTACCAACCAACTCGTACCAATCCAACTGATTCAATTGACA 1276
Db 1226 TATGCACTGGATCAGTACCACATCAAGTACTTATCAAAATCCCACTGACTCGGATGATA 1285
Qy 1277 CAGTGGTGTACAAAGTTCCACTGCCAAATCAAGTGTGTTAGTACTACTGAATATTTGGTCTC 1336
Db 1286 CAGTTGTTGTACAAAGTTCCACTGCCAAATCAACAGTTTAACTACTCAGTTTGGTTCAG 1345
Qy 1337 AGTCCTTTGTGTACAAACCACTACAGTTACTGCTCTCCAGGTGGTACCGACTCTGATTA 1396
Db 1346 GAAGTGTGCCCAACCAACCAACTGTGACCCTGACCAACCAAGGACGATAGTGTGATCA 1405
Qy 1397 TCAGAGACCAACCAACCACTACTGCTACTACTGTAATATTTGGTCAAAATCCCTTTGCTA 1456
Db 1406 TCAAGGACCAACCAACCACTACTGCTACTGCTACTGCTAGTTTGGTCAAAATCATTTGCTA 1465
Qy 1457 CTACTACTGTTACTGCTCTCCAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1516
Db 1466 CTACTGAGACAGTCCCAACCAATCCCAAGGCACTGATGTTGATCATCAAGGAACCA 1525
Qy 1517 CAAATCCAACCTGCTACTACCAACCGAGTATTTGGTCTCAATCCTTTGCTACTACTACTAC 1576
Db 1526 ACAATCCTACTGTTTACCAACCGAGTTTGGTCAAGTATTTGCTACTACTGAGACAG 1585
Qy 1577 TTACTGCTCTCCAGGTGTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1636
Db 1586 TCACCAACTATCCAGAGGGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1645
Qy 1637 TCACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1696
Db 1646 TAACAACCAAGGTTTGGTCAAGTATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1705
Qy 1697 CAGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1756
Db 1706 CAGAGGGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1765
Qy 1757 AATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1816
Db 1766 AGTTTGGTCAAGTATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1825
Qy 1817 ACAGTGTATCATTTAGAGACCAACCAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1876
Db 1826 ACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1885
Qy 1877 AATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1936
Db 1886 AATCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1945
Qy 1937 TTAGGGAACCA 1947
Db 1946 TACATGATCCA 1956

RESULT 12

AF201684

LOCUS

DEFINITION

complete cds.

ACCESSION

AF201684

VERSION

AF201684.1

KEYWORDS

SOURCE

ORGANISM

Candida albicans.

Candida albicans

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; mitosporic Saccharomycetales; Candida.

AF201684 Candida albicans agglutinin-like protein Als7p (ALS7) gene,
complete cds.

AF201684.1 GI:9754770

Candida albicans.

Candida albicans

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; mitosporic Saccharomycetales; Candida.

1 (bases 1 to 6897)

Hoyer, L.L. and Hecht, J.E.

The ALS6 and ALS7 genes of Candida albicans

AF201684 6897 bp DNA linear PLN 09-AUG-2000


```
Qy 964 TCTAAGGATATGTCATGTTGCTACAACTAGAACAGTTACAGACAGTACCACTGCTGTC 1023
Db 970 GCTATTGGTCAGACCTAGTTTCGTACCACTCCACAGGATAGTTTCATTACCAAGAACT 1029
Qy 1024 ACTACTTTTACCATTCAATCCAAAGTGTGATATAAACCACCAAAACAAATCGAAATTTTGCACCT 1083
Db 1030 ACCACATTACCCCTTCATTTCCCGACTCCAGAAACCAAAACAAATTCCTAGTCTTTAGAGCCC 1089
Qy 1084 ATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCTATCTGACTAAG 1143
Db 1090 ATACCCCACTACCGTAAACAACITTCACACATGCTTTGATACTTGGTATATATACPAAG 1149
Qy 1144 ACTGCACCAATTTGGTGAACAGCTACTGTTATTTGTTGATGTCATATCATATCATATACACACA 1203
Db 1150 AAAGCCACCAATTTGGTGACACAGCTACTGTTTTCATTTGATGTTCCACACATACAGCTACT 1209
Qy 1204 ACTGTTACCAAGTGTGACAGGAGCAATCACTACCAACCAACTCGTACCAATCCCACT 1263
Db 1210 ACTTTGACCAATATTTGGCAAGAATCAAGTACAGGACCAACCACTTACTTCGATGACATA 1269
Qy 1264 GATTCAATTTGACACAGTGGTGTACAGTTCCCACTGCCAAATCCAACTGTTAGTACTACT 1323
Db 1270 GACTTGGTCGATACTGTCATTTGTGAATAATCCATATCCCAATCCGACTATTATTAACAACA 1329
Qy 1324 GAATATTGGTCTCAGTCTCTTTGCTACAACTACAGTACTGCTCTCCAGGTGGTACC 1383
Db 1330 CAATTTTGGTCAGGTAATATTTAACTACTGAGACACACAAAGAACCACTCTCGGTACT 1389
Qy 1384 GATACTGTGATTATCAGAGAGCCACCAACCACTACTGTCATCTACTGATGAATATTGGTCA 1443
Db 1390 GATAGTGTGATCATCAAGGAACCAACCACTACTGTGACAAACGACGAGTTTGGTCA 1449
Qy 1444 CAATCTTGTCTACTACTACTGTTTACTGCTCTCCTCCAGGTGGTACTGACTCAGTAAT 1503
Db 1450 GAATCATTTGCCACTACTGAGACCACTACCAACTATCCAGAAGGCAGTACTGACTCAGTCATT 1509
Qy 1504 ATCAGAGAACCAACCAATCCAACTGTCATACAACTGTCATACAGCGAGTATTTGGTCTCACTCTTGGCT 1563
Db 1510 GTTAGGGAACACACAACTGTCATGTCAGACGACGAGTTTGGTCAAGATCAATTTGCC 1569
Qy 1564 ACTACTACTACAGTTACTGCTCTCCAGGTGGTACTGACTCAGTAATTTATCAGAGAACCT 1623
Db 1570 ACTACTGAGACCACTACCAACCGGTCAGAGGCAGTACTGACTCAGTCATTGTTAGGGAACCA 1629
Qy 1624 CCAACCACTGTCACCACTGATATTTGGTCCCAATTTAGGTCACCACTGTAACCTGACCA 1683
Db 1630 CATATCAACTGTGACAAACCGAGTTTGGTCAAGATCAATTTGCCACTACTGAGACCACTGAGACC 1689
Qy 1684 GTGACTGCTCTCCAGGAGGCACTGACTCAGTAATTTATCAGAGAACCAACCACT 1743
Db 1690 ATCAACAGGTCAGAGAGGCACTGACTCAGTCATTGTTAGGGAACCAACCAATCCCACT 1749
Qy 1744 GTCACTACTGTAATCTGTCACAACTCATATGTCACCACTACCACTGTAACCTGACCA 1803
Db 1750 GTGACAAACCGAGTTTGGTCAAGATCAATTTGCCACTACTGAGACCACTACCAACCGT 1809
Qy 1804 CCAGTGGTACTGACACTGTTATCATTTAGAGAGCAACCAACCACTGTCACCTACTACT 1863
Db 1810 CCAGAAGCACTGACTCAGTCATTGTTAGGGAACCACTAATCCAACTGTGACAAACCA 1869
Qy 1864 GAGTATTGGTCTCAATCTTTGCTACTACCACTGTAACCTGGTCCACCAAGTGGCACT 1923
Db 1870 GAGTTTGGTCAGATCAATTTTGGCACTACTGAGACCACTACCAACCGGTCAGAGAGGCACT 1929
Qy 1924 GATAGTGTATCATTTAGGGAACCAACCACTGTCCACTACTGTAATCACTGGTCT 1983
Db 1930 GATAGTGTATCATCAAGGAACCAACCACTGTGACTACTGACTACCAACAGTTTGGTCA 1989
Qy 1984 CAATCATATGCAACCACTACTACCAATACCCTCCACTGTTGATGAACTGATACCGTTCT 2043
Db 1990 GAATCATTTGGCACTACTGAGACCAATACCAACTATCCAGAAGGCACTGACTCAGTCATT 2049
Qy 2044 ATCAGAGAGCCACCAACCACTACTGTCATCTACTGATGATCTGCTCAATCATATGCT 2103
```

```
Db 2050 GTTAGGGAACCACTAATCCAACTGTGACAAACCCAGTTTGGTCAGAATCATTTGCC 2109
Qy 2104 ACAACCACTGTTTACTGTCACCACTGGTGAACCCGATACCGTCTCTTATCAGAGACCA 2163
Db 2110 ACTACTGAGACCATCACCAACGGTCCAGAGGCACTGACTCAGTCATTGTTAGGGAACCA 2169
Qy 2164 CCAACCACTGTCACCTACTACTGTAATCTGCTCAATCATATGCTTACCAACCACTACT 2223
Db 2170 CACAATCCAACTGTGACAAACCAACCGAGTTTGGTCAAGATCATTTGCTACTACTGAGACC 2229
Qy 2224 GTTACTGACCACTGTCAGGTGATACCACTGTTTATCATTTAGAGAGCCACCAATCCCAACA 2283
Db 2230 ATCAACCACTATCCAGAGGCACTGACTCAGTCATTGTTAGAGAACCAACCAATCCCAACT 2289
Qy 2284 GTTACTACTGTAATATTGTCACAACTTTGTCACAACTTTGTCACAACTTTGCTGCTCT 2343
Db 2290 GTACAAACCACTGTTTGGTCAAGTCATTGCTACTGAGACAGTACCACCACTAT 2349
Qy 2344 CCAGGTGCTACTGACACTGTTGATTATCTATGAA 2376
Db 2350 CCAGAAGCACTGACTCAGTCATTGTTAGGGA 2382

RESULT 13
CANALS4S1
LOCUS
DEFINITION
  Candida albicans agglutinin-like protein (ALS4) gene, 5' partial
  cds.
ACCESSION
  AF024584
VERSION
  AF024584.1 GI:3598677
KEYWORDS
  1 of 2
SEGMENT
  1
SOURCE
  Candida albicans.
ORGANISM
  Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
  Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE
  Hoyer,L.L., Payne,T.L. and Hecht,J.E.
AUTHORS
  Identification of Candida albicans ALS2 and ALS4 and localization
  of als proteins to the fungal cell surface
  J. Bacteriol. 180 (20), 5334-5343 (1998)
JOURNAL
  98440424
MEDLINE
  9765564
PUBMED
  2 (bases 1 to 1407)
REFERENCE
  Hoyer,L.L.
AUTHORS
  Direct Submission
  Submitted (11-sep-1997) Veterinary Pathobiology, University of
  Illinois, 2001 S. Lincoln Avenue, Urbana, IL 61802, USA
JOURNAL
  Location/Qualifiers
FEATURES
  source
  1..1407
    /organism="Candida albicans"
    /strain="1161"
    /db_xref="taxon:5476"
    /chromosome="6"
    /map="6C"
    1..>1407
    /gene="ALS4"
    /note="agglutinin-like sequence"
    /allele="ALS4-1"
    1..>1407
    /gene="ALS4"
    /codon_start=1
    /transl_table=12
    /product="agglutinin-like protein"
    /protein_id="AAC64239.1"
    /db_xref="GI:3598680"
    /translation="MLLQFLLSLCVSVATAKVITGVFNSFNLSLTWANAASYPYRGA
    PTFWTAVIGNSLDGATASAGDTFTLDMPCVFKFTIDQTSIDLVDGRTYATCNLSAE
    EPTFSSVSCVTVTMTADTKAIGTVPFSPVSGSGSDVDLANSQCFTAGINTVF
    NQDGTSTSTVDFEKSTVASSDRILLRILPSLQAVNLFLPQCANGYTGTMGFT
    AGTCATIDCSTVHVGIENGLNDWNPISSEFSFVTKTCTSTSVLVTQNPVAGYRPFV
    DAYISATRVSSYTMQYTNIVACGAASVDDSFTHWRGYSNQSAGSNGITIVTTRTV
```

TDSTAVTLFPNSDIDKTKIEILOPIPTTTITTSYVGVTSYKTPAPIGETATVI
VDVYHTTTTTSWTGTTTITTTTNPDSIDIVVQVSPNPVTVTTEVWSQSYAT
TTTTVAPGGTDSVIIEP"
1300..1407

repeat_region

/note="first of multiple copies of 108bp imperfect repeat"

/rpt_type=tandem

BASE COUNT 391 a 294 c 253 g 469 t

ORIGIN

Query Match 22.3%; Score 842.8; DB 8; Length 1407;
Best Local Similarity 75.0%; Pred. No. 2.8e-142;
Matches 1054; Conservative 0; Mismatches 352; Indels 0; Gaps 0;
QY 1 ATGCTTCAACAATTTACATTTGTTTCCATATTTGTTCAATGCAAGTGCACAAAGACAATC 60
DB 1 ATGCTTTTACAATTTTGTGTAAGCCCTCTGTATATCATGTTGCTACGGCAAAAGTTAT 60
QY 61 ACTGGTGTGTTTGTAGTATTTTAAATCAATTAATGTTGTTCAATGCTGCTAAATATGCTTTTC 120
DB 61 ACAGGTGTGTTTCAATAGTATTTTAAATCGTTAACTTGGGCCAATGCTGCTTCTTTATCCATAT 120
QY 121 AAAGGCCAGGATACCCCAACTTGGAAATGCTGTTTGGGTTGCTCTAGATGTTACCCAGT 180
DB 121 AGAGGTCCAGTACTCCTACTTGGACCGCTGTAAAGGATGGTCTTTAGATGGAGTACT 180
QY 181 GCCAATCCAGGGGATACATTCACATTTGAATATGCCATGTGTGTTTAAATATATACTACTTCA 240
DB 181 GCTAGTGTGTTGGACACATTCACGTTAGACATGCTTGTGTTTCAAAATTTATTACTGAT 240
QY 241 CAACATCTGTTGATTTTAACTCCGATGTTTAAATATGCTGCTACTGTGCTCAATTTTATCT 300
DB 241 CAAAGCTCAATTTGATTTAGTGTGCTGCTGCTACTTGTGCTGCTGCTGCTGCTGCTGCTGCT 300
QY 301 GGTGAAGAAATTCACAACTTTTCTACATTAACATGTACTGTGCAAGCAGCTTTGAAATCA 360
DB 301 GCCGAAGAGTTTACTACTTTTCTAGTGTGCTATGCTACTGCTACTACTACATGACTGCT 360
QY 361 TCCATTAAAGGCATTTGGTACATTTACTTTTACCAATTTGCATTCATCAATTTGGTGGACAGGT 420
DB 361 GACACCAAGCCATAGGAATCTTAACATTTACCTTTCTCATTCAGTGTGGGGGATCAGT 420
QY 421 TCATCACTGATTTGGAAGATTTCTAAATGTTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
DB 421 TCAGATGTTGATTTGGCAATTTCTCAATGTTTACTGTCAGGAATCAATACAGTTTACTTTT 480
QY 481 AATGATGTTGATAAGATATCTCAATTTGATGTTGAGTTTGAAGTCAACCGTTGATCA 540
DB 481 AATGATGTTGACACTAGCATTTCCACACAGTTGATTTTGAATAATCAACCGTGGCTCC 540
QY 541 ATGTCATATTTGATGCTTCCAGAGTTATGCCAAGTCTCAATTAAGTCAACAATCTTTT 600
DB 541 AGCGATCGTATCTTGTGTCAAGATTTTACCAGTCTTTTCAACAGCAATAATCTTTT 600
QY 601 GTGGCACCACATGCAAAATGGTTACACATCTGTTACATGCTGGTACATGGGTTCTCCAGTAGTAC 660
DB 601 CTTCGCCAAGATGCAAAATGGTTATCTTCTGTTACAAATGGGATTTTTCGACTGCTGT 660
QY 661 GGTGAGTGTGCTATTGCTCAATATTTTCATATTTGTTGATGATCACAAAGGATTAATATGAT 720
DB 661 ACTGGTGTCTATAGATTTTCCAGTTCATGTCGGGATATCAAAATGGTGGTGAATGAT 720
QY 721 TGAATTTATCCGGTTTCACTGCTGAATCATTTAGTTACACTTAAACCTTTGATCTCAATGGA 780
DB 721 TGAATTTATCCAAATTTTCACTGGAATCTTTTCTTACACAAAGACCTGTATACATCAACAAGT 780
QY 781 ATTACAGATTAATATCAAAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
DB 781 GTTTTAGTAACTTTTCAAAATGTTCTGCGGATATGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
QY 841 TCTGCTACAGATGTTAAACCAATATACTTTAGCATATACCAATGATTAATCTTGTGCTGCTG 900
DB 841 TCTGCCACACAGTACGCTCATATACCATGCAATACACTAATATATATATGCTTGTGCTGCTG 900

QY 901 AGTCGTCTCAAAAGTAAACCTTTTCACTTTAAGATGGACTGGATACAAAGAAATAGTGATGCC 960
DB 901 GCGGCTCTCTGTGATGACTCACTTTACTCATCTTGGCGGGATATAGTATAGTCAAGCT 960
QY 961 GGATCTAACCGTATTGTTCACTTTGTTGCTCAACAATAGAACAGTTTACAGACAGTACCAGTCT 1020
DB 961 GGTTCCTAATGTTATTACCAATTTGTTGTTAACAATAGAACAGTTTACAGACAGTACCAGTCT 1020
QY 1021 GTCACTACTTTTACCAATTCATCAATCAAGTGTGATAAACCAAAACAAATCGAAATTTTGCAA 1080
DB 1021 GTGACTACTTTTACCAATTTCAATTTCCGATACTGACAAACCAAAACAAATCGAAATTTTACAA 1080
QY 1081 CCTATTCCAAACCACTACCACTACCACTCATATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1140
DB 1081 CCTATTCCAAACCACTACCACTACCACTCATATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1140
QY 1141 AGACTGCAACCAATTTGGTGAACAGTACTGTTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1200
DB 1141 AAACTGCAACCAATTTGGTGAACAGTACTGTTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1200
QY 1201 ACAACTGTTTACCAGTGAATGGACAGCAACAATCACTACCACCACTCTGTACCACCAATCCA 1260
DB 1201 ACAACTGTTTACCAGTGAATGGACAGCAACAATCACTACCACCACTCTGTACCACCAATCCA 1260
QY 1261 ACTGATTCATTTGACACAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1320
DB 1261 ACTGATTCATTTGACACAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1320
QY 1321 ACTGATATTTGTTCTCAGTCCCTTTGCTACCAACCACTACAGTACTGCTCTCCAGTGGT 1380
DB 1321 ACCGAGTATTTGTTCTCAGTCCCTTTGCTACCAACCACTACAGTACTGCTCTCCAGTGGT 1380
QY 1381 ACCGACTACTGTTGATTTATCAGAGAGCC 1406
DB 1381 ACTGACTCAGTAATTTATCAGAGAGCC 1406

RESULT 14

CANALS42S1

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SEGMENT

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

gene

1 of 2

Candida albicans.

Candida albicans.

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; mitosporic Saccharomycetales; Candida.

1 (bases 1 to 1407)

Hoyer, L.L., Payne, T.L. and Hecht, J.E.

Identification of Candida albicans ALS2 and ALS4 and localization

of als proteins to the fungal cell surface

J. Bacteriol. 180 (20), 5334-5343 (1998)

9840424

9765564

2 (bases 1 to 1407)

Hoyer, L.L.

Direct Submission

Submitted (11-SEP-1997) Veterinary Pathobiology, University of

Illinois, 2001 S. Lincoln Avenue, Urbana, IL 61802, USA

Location/Qualifiers

1. .1407

/organism="Candida albicans"

/strain="1161"

/db_xref="taxon:5476"

/chromosome="6"

/map="6C"

1. >1407

/gene="ALS4"

/note="agglutinin-like sequence"

/allele="ALS4-2"

```

COS
1. >1407
/gene="ALS4"
/codon_start=1
/transl_table=12
/product="agglutinin-like protein"
/protein_id="AAC64241.1"
/db_xref="GI:3598685"
/translation="MLIQFLLSLCSVSATKAVITGVNSFNSLWANAASYPYRGA
EPTPTWIGSLDCAATAGKDTFLDMPCKFKITDOTSIDLVADGRTYATCNLSGAE
EFTFSVSCVTMTTADTKAMGTLPFSPFVGGSGSDVLANSQCFTAGLNTVTF
NDGDTISITVDEKSTVASSDRILLRILPSLSQAVSLPQECANGTSTGTMGRST
AGTACGCTSVHVGSLNGLDNWNYPIDSEFSYTKTCTSTSVLYTVQNPVPGYRPFV
DAYSATRVSSYTMQNTIYACGAASDVSEFTHWRGYSNQAQNGNITIVYTRT
TDSSTAVTLTPNSDDKTKIEILQIPITPTTITTSYVGVTTSYSTKTAIPIGTATVI
VDVYPTTLTTESEMTGITVTITTTTRTNPTDSIDTVVQVPSNPPTVTITTEYWSQSYAT
TTVTAPPNGDTSVILREP"
repeat_region
1300..1407
/note="first of multiple copies of 108bp imperfect repeat"
BASE COUNT 389 a 297 c 257 g 464 t
ORIGIN
Query Match 22.1%; Score 836.4; DB 8; Length 1407;
Best Local Similarity 74.7%; Pred. No. 4e-143;
Matches 1050; Conservative 0; Mismatches 356; Indels 0; Gaps 0;
QY 1 ATGCTTCAACAATTTACATCTGTTATTCCTATATTTGTCAAATGCAAGTGCACAAACAATC 60
DB 1 ATGCTTTTACAAATTTTGTGCTAAGCCTCTGCTATCAGTTGCTACGGCAAAAGTTATT 60
QY 61 ACTGGTCTTTTGTATAGTTTAAATTCATTAACCTGGTCCAAATGCTGCTAAATATGCTTTC 120
DB 61 ACAGGTCTTTTCAATAGTTTAAATTCGTTAACTGGGCCAATGCTGCTCTTATCCATAT 120
QY 121 AAAGGGCCAGGATACCAACTTGAATGCTGTTTGGTGGTCCCTTAGATGGTACCAGT 180
DB 121 AGAGGTCACGCTACTCCTACTTGGACGCGCTGTAATAGTAGTGGTCTTTAGATGGAGCT 180
QY 181 GCCAATCCAGGGATACATTCACATTCGAATATGCCATGCTGTGTTTAAATATACACTTCA 240
DB 181 GCTAGTCTGGTGACATTCACGTTAGACATGCGTTGTGTTTCAAAATTTATTACTGAT 240
QY 241 CAAACATCTGTGATTTAACTGCCGATGGTGTAAATATGCTACTTGTCAATTTTATTTCT 300
DB 241 CAAACGTCAAATGATTTAGTTGCTGATGGTGTACTTATGCTACTTGTAAATTTCAACTCT 300
QY 301 GGTGAAGAATTCACAACTTTTCTACATTAACATGTACTGTGACGACGCTTTCGAATCA 360
DB 301 CGCGAAGAGTTTACTACTTTTCTAGTGTGTCATGTTACTGTGACTACTACAAATGACTGCT 360
QY 361 TCCATTAAAGCATTTGCTACAGTTTACTTTTACCAATTCGATTCATTTGGTGGGAACAGT 420
DB 361 GACACCAAGCCATGGGAAGTGAACATTAACCTTCTCATTCAGTGTGGGGGATCAGT 420
QY 421 TCATCAACTGATTTGGGAAGATCTTAAATGTTTTACTGCTGGTACCAATACAGTCACAFTT 480
DB 421 TCAGATCTGATTTGGCAAAATCTCAATGTTTTTACTGCAGGAATCAATACAGTTTACTTTT 480
QY 481 AATGATGGTGATAAAGATATCTCAATTTGATGTTGAGTTTGAAGAACTCAACGTTGATCCA 540
DB 481 AATGATGGTGACACTAGCATTTTCCACAACTGATGATTTTGAAGAAATCAACCGTGGCCTCC 540
QY 541 AGTCATATTTGATGCTCCAGAGTTTATGCCAAGTCTCAATAAGGTACACAACCTCTTTT 600
DB 541 AGCGATCGTATCTGTTGTCAGAAGATTTTACCCAGTCTTTCACAAGCAGTAAGTCTTTTT 600
QY 601 GTGSCACCACAATGTGAAAATGGTTACACATCTGGTACAATCTGGGGTCTCCAGTAGTAAC 660
DB 601 CTTCCCCAAGAATCTGCAAAATGGTTATCTCTGTCACAATGGGATTTTCAACTGCTGGT 660
QY 661 GGTGACGCTTGCTATGATGCTCAAAATATTCATATTCGATATCACAAAAGGATTAATGAT 720
DB 661 ACTGGTGCTACTATAGATTTGTTCCACAGTTCATGCTCGGGATATCAAAATGGGTGCAATGAT 720

```

```
/organism="Candida albicans"
/strain="1161"
/db_xref="taxon:5476"
/chromosome="6"
/map="60"
1. >1404
/gene="ALS9"
/codon_start=1
/transl_table=12
/product="agglutinin-like protein"
/protein_id="AAK00764.1"
/db_xref="GI:12656146"
/translation="MLPQIFLLISLTAKTIRGVNFSFSLTWRSVEYAYKGE
PTFWAVLWLSNSTADPGDTFTLILCVFKFITQTSVDLTADGVSYCDFNAGE
EFTFSSLSCTVNSVSYSVARGVKLPITFNVGKTSSVDLADSKCFTAGKNVTF
MDGNPKISTVDVDFDASPVSPSGYITSSRIISLNLKLSLFLVPOCENGYSGIMGFVA
SNGAIDSNKYLITGLNDWNPVSSSESYTKTSTISITVFQNPVAGPRFVD
ATISENDKNTLYITANEYTCENGNTVDPFLLIWMGYKNSDSDGVIVVVTITVT
DSTTAVTLTPNPSVDKTEITELQPIPTTTTTSYIGISTSYETLGTIGGTATVIV
DTPYHITATVNFMTGSGITTTTNTPTGSDIVIVQIPSPDPTTTITFEFWSSEFAST
TTITNPDPGTSNVIIEKP"
-repeat_region
1297..>1404
/notes="first of a series of 108 bp repeat units"
/rpt_type=tandem
413 a 286 c 452 t
```

```
BASE COUNT 413 a 286 c 452 t
ORIGIN
Query Match 21.3%; Score 805.4; DB 8; Length 1404;
Best Local Similarity 73.8%; Pred. No. 1.6e-135;
Matches 1038; Conservative 0; Mismatches 366; Indels 3; Gaps 1;
QY 1 ATGCTTCAACAATTACATTTGTTATTCCTATATTTGTCATTTGCAATGCAAGTGCAGAACACAATC 60
Db 1 ATGCTTCCACAATTCATATTTGTTATTCATATCTTGCACAGTGCAGTGCAGTGCAGAACTATTT 60
QY 61 ACTGGTGTGTTGATAGTTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTT 120
Db 61 ACTGGTGTGTTGATAGTTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTT 120
QY 121 AAAGGCCAGGATACCACTTGGATGCTGTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 180
Db 121 AAAGGCCAGGATACCACTTGGATGCTGTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 180
QY 181 GCAATCCAGGGGATACATTCACATTCATATGCAATGCAATGCAATGCAATGCAATGCAATGCAATTC 240
Db 181 GCTGACCCAGGAGACATTCACATTCATATGCAATGCAATGCAATGCAATGCAATGCAATGCAATTC 240
QY 241 CAAACATCTGTTGATTAAGTCCGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 300
Db 241 CAAACATCTGTTGATTAAGTCCGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 300
QY 301 GGTGAAGAATTCACAACTTTTCTACATTAACATGACTGTAACGACGCTTTTGAATATCA 360
Db 301 GGTGAAGAATTCACAACTTTTCTACATTAACATGACTGTAACGACGCTTTTGAATATCA 360
QY 361 TCATTTAAGGCATTTGGTACAGTACTTTACCAATTCATTCATTCATTCATTCATTCATTCATTCATTC 420
Db 361 TATGCTAGGTTTCTGTCAGGTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 420
QY 421 TCATCAACTGATTTGGAAGATCTTAATGTTTACTGCTGGTACCATACAGTACATTCATTCATTCATTC 480
Db 421 TCATCAACTGATTTGGAAGATCTTAATGTTTACTGCTGGTACCATACAGTACATTCATTCATTCATTC 480
QY 481 AATGATGGTGATTAAGATATCTCAATGATGTTGAGTTTGAAGTCAACCGTTGATCCA 540
Db 481 ATGGATGGCGGATCAAGATATTTTACCACTGTTGATTTGACCGCTCTCCAGTATCACCC 540
QY 541 AGTGCATATTTGTTATGCTTCCAGATTTATGCCAAGTCTCAATAGGTGCAACACTCTTTT 600
Db 541 AGTGGTTATATACAGCTCACCAATATTTCCATAGTCTCAATAACTTTCAAGTCTTTTC 600
QY 601 GTGGCACCAATGTGAAATGGTTACACATCTGGTACATGGGTTCTCCAGTAGTAAC 660
Db 601 GTGGCACCAATGTGAAATGGTTACACATCTGGTACATGGGTTCTCCAGTAGTAAC 660
```

```
Db 601 GTGGTCCCAATGTGAGAATGTTACACATCTGGTATAATGGATTTGTAGCTAGTAAC 660
QY 661 GGTGACGTTGCTATTCCTCAATATTCATATTTGGTATACAAAGGATTAATATGAT 720
Db 661 GGT---GCTACTATTTGCTCAATGCAATATAGGAATATCAAAAGGTTTAAATGAT 717
QY 721 TGAATATTCGGGTTTCATCTGAATCATTTAGTTACACTAAAACTTTGTACATCTAATGGA 780
Db 718 TGAATTTTCCAGTAAGTTTCAGAAATCATTTCTTACACGAAAACTTTGTAGTCAACCACT 777
QY 781 ATTCAGATTAATATCAAAATGTACCTGCTGGTTATCGTCCATTTATTTGATGCTTATAT 840
Db 778 ATTCAGATTTGAATTTCAAAATGTCTGCTGGTATCGCCCTTTTGTGATGATATAT 837
QY 841 TCTGCTACAGATCTTAAACCAATATCTTTAGCATATACCAATCATATCTTCTGCTGCGC 900
Db 838 TCTGCGAGAAATATTCGTAATATATCTTGCACGTACGCAATAGTATATCTTCTGGAAT 897
QY 901 AGTCTGCTGCAAAAGTAAACCTTTTCACTTTAAGATGACCTGGATACAGAAATAGTATGCC 960
Db 898 GGCATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 957
QY 961 GGATCTAAACGGTATTTGCTCATTTGCTACAACTAGAACAGTTACACAGTACCACTGCT 1020
Db 958 GACTCTGACGGGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1017
QY 1021 GTCACCTACTTTACCATTTCAATCCAACTGTTGATATAAACCAAAACAATCGAAATTTTGCAA 1080
Db 1018 GTGACTACTTTTACCTTTCAATCCAACTGCTGATATAAACCAAAACAATTTGAATTTTGCAA 1077
QY 1081 CCTATTCCAACTACTACCATTCACAACTTCATATGTTGGTGTGACTTCTCTCTCTGACT 1140
Db 1078 CCATTTCCCAAGCAGCACCAATTCACAACTTCATATATTTGGTATTTTCCACTTCTCTATGAACA 1137
QY 1141 AGACTGCAACCAATTTGGTGAACAGCTACTGTTATTTGTTGATGTGCTATCATCTACTACC 1200
Db 1138 TTAACCGGAACAATTTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1197
QY 1201 ACAACTGTTTACCACTGCAATGGACGGAACAATCACTACCAACCACTGCTACCAATCCA 1260
Db 1198 GGCACCTGTTACAAATTTCTGGACTGGTGCANTTACAACTACCACTACTTATCTAATCCC 1257
QY 1261 ACTGATTTCAATTCACACAGTGGTGGTACAAAGTTCCACTGCCAAATCCAATGTTAGTACT 1320
Db 1258 ACTGGTTCCATAGACACTGTTATTTGCAAAATTTCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1317
QY 1321 ACTGAATATTTGGTCTCAGTCTTGTGCTACAAACCACTACAGTTACTGCTCTCTCCAGGTGGT 1380
Db 1318 ACTGAATTTTGGTCTGAACTATTTGCTAGTACTACCACTACCACTACCACTACCACTGAGGGT 1377
QY 1381 ACCGATACTGCTGATTTACAGAGGCCA 1407
Db 1378 ACTAATAGTGTGATCATCAAGAACCA 1404
```

Search completed: April 29, 2003, 19:18:22
Job time : 6516 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search; using sw model

Run on: April 29, 2003, 15:36:18 ; Search time 516 Seconds
(without alignments)
16523.383 Million cell updates/sec

Title: US-09-715-876-7
Perfect score: 3786
Sequence: 1 atgcttcacaaattacatt.....tgtcttgcatttagtga 3786

Scoring table: IDENTITY_NUC
Gap 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues 4370478
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	359.2	9.5	424	17 AAT29067	Probe for Candida
2	108	2.9	108	17 AAT32337	Probe for Candida
3	108	2.9	108	17 AAT32330	Probe for Candida
4	108	2.9	108	17 AAT32331	Probe for Candida
5	108	2.9	108	17 AAT32332	Probe for Candida
6	108	2.9	108	17 AAT32334	Probe for Candida
7	108	2.9	108	17 AAT32336	Probe for Candida
8	108	2.9	108	17 AAT29063	Probe for Candida
9	106.4	2.8	108	17 AAT32333	Probe for Candida

10	106.4	2.8	108	17	AAT32335	Probe for Candida
11	99	2.6	99	17	AAT32338	Probe for Candida
12	94.2	2.5	8201	21	AA88864	Human dentin sialo
13	94.2	2.5	8201	24	ABQ73537	Human dentin sialo
14	87	2.3	5511	21	AA61847	Cryptosporidium pa
15	87	2.3	7334	21	ABT04776	C parvum Gp900 gen
16	87	2.3	7334	21	AA61846	Cryptosporidium pa
17	87	2.3	7334	24	ABT04775	C parvum Gp900 gen
18	86.2	2.3	5163	19	AAV20700	Cryptosporidium pa
19	86.2	2.3	5163	21	AA61849	ORF encoding a por
20	86.2	2.3	5163	24	ABT04778	C parvum Gp900 gen
21	86.2	2.3	5318	19	AAV20701	Cryptosporidium pa
22	86.2	2.3	5318	21	AA61848	DNA encoding a por
23	86.2	2.3	5318	24	ABT04777	C parvum Gp900 gen
24	85.2	2.3	108	17	AAT29064	Probe for Candida
25	82.4	2.2	100	15	AAQ62588	Candida albicans-s
26	82.4	2.2	100	15	AAQ62589	Candida albicans-s
27	81.4	2.2	7758	24	ABL33102	Human immune syste
28	76	2.0	100	15	AAQ62589	Candida albicans-s
29	74.4	2.0	267156	24	ABL68560	Kidney cancer rela
30	68.4	1.8	40875	18	AAT80043	Insert from cosmid
31	68	1.8	100	15	AAQ62592	Candida albicans-s
32	68	1.8	7758	24	ABL33103	Human immune syste
33	67.6	1.8	2336	23	ABL25662	Drosophila melanog
34	64.8	1.7	2790	7	AA60473	Sequence encoding
35	62.6	1.7	1368	10	AA91235	DNA sequence of pr
36	62.2	1.6	2014	24	ABA90791	Bacillus anthracis
37	58	1.5	3168	20	AA36552	C. albicans Rbt1 c
38	58	1.5	3983	23	ABL15838	Drosophila melanog
39	57.4	1.5	2335	23	ABL25476	Drosophila melanog
40	57.2	1.5	100	15	AAQ62593	Candida albicans-s
41	57.2	1.5	3399	17	AAT05868	Chicken leucocytoz
42	57.2	1.5	32392	24	ABL56203	AmEPV genome fragm
43	57.2	1.5	50000	24	ABL5643	AmEPV genome fragm
44	57	1.5	100	15	AAQ62591	Candida albicans-s
45	56.6	1.5	14987	24	ABL32630	Human immune syste

ALIGNMENTS

RESULT 1
AAT29067
ID AAT29067 standard; DNA; 424 BP.

AC AAT29067;

XX 28-NOV-1996 (first entry)

XX Probe for Candida albicans and Candida stellatoideis.

DE Probe: primer; detection; identification; Candida albicans;

KW Candida stellatoideis; sputum; bronchial washings; blood; milk;

KW lymph fluid; skin; soft tissue; ss.

XX Synthetic.

XX WO9618745-A1.

XX 20-JUN-1996.

XX 08-DEC-1995; 95WO-US16153.

XX 16-DEC-1994; 94US-0357962.

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX Hoyer LL, Livi GP, Shatzman A;

XX WPI; 1996-300661/30.

XX C. albicans and C. stellatoideis specific probes and primers - for

PT specific detection of Candida infection


```

DT 29-NOV-1996 (first entry)
XX Probe for Candida albicans and Candida stellatoideis.
DE
XX
XX Probe: primer: detection; identification: Candida albicans;
KW Candida stellatoideis; sputum; bronchial washings; blood; milk;
KW lymph fluid; skin; soft tissue; ss.
XX Synthetic.
OS
XX WO9618745-A1.
PN
XX 20-JUN-1996.
PD
XX 08-DEC-1995; 95WO-US16153.
PF
XX 16-DEC-1994; 94US-0357962.
PR
XX (SMIK ) SMITHKLINE BEECHAM CORP.
PA
XX Hoyer LL, Livi GP, Shatzman A;
PI
XX WPI; 1996-300661/30.
DR
XX C. albicans and C. stellatoideis specific probes and primers - for
PT specific detection of Candida infection
PT
XX Claim 1; Figure 5A; 33pp; English.
PS
XX Five synthetic sequences (AAT29063-67) are used as probes to detect
CC the presence of Candida albicans and Candida stellatoideis in
CC biological fluids e.g. sputum, bronchial washings, blood, milk and
CC lymph fluid or in tissue samples e.g. skin and soft tissues. The
CC method of detection may also comprise using two of the probe
CC sequences (AAT29065, AAT29066) as amplification primers and then
CC contacting one of the probe sequences with the amplified product and
CC detecting hybridisation.
XX
XX Sequence 99 BP; 29 A; 29 C; 14 G; 27 T; 0 other;
SQ
Query Match 2.6%; Score 99; DB 17; Length 99;
Best Local Similarity 100.0%; Pred. No. 2.7e-13;
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2272 CCAATCCACAGTTACTACTGTAATATGTCACAAATATGTCACAAACACCA 2331
Db 1 CCAATCCACAGTTACTACTGTAATATGTCACAAATATGTCACAAACACCA 60

Qy 2332 GTTACTGCTCCTCCAGGTGGTACTGACACTGTGATTATC 2370
Db 61 GTTACTGCTCCTCCAGGTGGTACTGACACTGTGATTATC 99

RESULT 12
AA88864/C
ID AA88864 standard; DNA; 8201 BP.
XX
AC AA88864;
XX
XX 19-FEB-2001 (first entry)
DT
XX Human dentin stialophosphoprotein gene.
DE
XX Dentin stialophosphoprotein; DSPP; human; SIBLINGS;
KW integrin-binding ligand; diagnosis; antiinflammatory; therapy;
KW osteoporosis; chromosome 4; ds.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH 2387..7896
FT CDS /*tag= a
FT /*note= "contains exons"

```

```

FT exon 2387..2437
FT /*tag= b
FT intron 2438..3576
FT /*tag= c
FT exon 3577..3660
FT /*tag= d
FT intron 3661..3793
FT /*tag= e
FT exon 3794..4780
FT /*tag= f
FT intron 4781..5256
FT /*tag= g
FT exon 5257..7896
FT /*tag= h
XX
XX WO200062065-A1.
PN
XX 19-OCT-2000.
PD
XX
XX 07-APR-2000; 2000WO-US09349.
PF
XX
XX 09-APR-1999; 99US-0128468.
PR
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA
XX Fisher LW, Fedarko NS, Young MF;
PI
XX WPI; 2000-679515/66.
DR
XX P-PSDB; AAB19772.
DR
XX Detecting small integrin-binding ligand N-linked glycoproteins for
PT detection of a tumor or protection against a complement mediated immune
PT response, comprises detection where Factor H is not an inhibitor -
XX
XX Disclosure; Page 98-100; 110pp; English.
XX
XX The present sequence is that of DNA encoding human dentin
CC stialophosphoprotein (DSPP, see AAB19772), a member of the small
CC integrin binding ligand, N-linked glycoproteins (SIBLINGS) family.
CC The invention provides methods and compositions for exploiting the
CC discovery that members of the SIBLINGS family bind to complement
CC Factor H, conferring resistance to complement mediated lysis. A
CC claimed method of conferring protection against a complement
CC mediated immune response involves providing a reservoir or other
CC supply in the subject's body so that a SIBLINGS protein can be
CC dispersed to interfere with complement mediated lysis and
CC inflammation. This protects cells that are grafted onto foreign
CC tissue or bone marrow cells introduced into a foreign host. The
CC SIBLINGS protein can be BSP, OPN, DMP1 or DSPP. A method of
CC detecting a SIBLINGS protein in a sample from a subject suspected
CC of having abnormal bone turnover, especially osteoporosis, is also
CC claimed.
XX
XX Sequence 8201 BP; 2801 A; 1477 C; 1873 G; 2050 T; 0 other;
SQ
Query Match 2.5%; Score 94.2; DB 21; Length 8201;
Best Local Similarity 41.7%; Pred. No. 1.6e-11;
Matches 588; Conservative 0; Mismatches 823; Indels 0; Gaps 0;

Qy 952 AGTCATCGCGATCTAACGGTATTGTCATTGCTCTACACTAGACAGTTACAGACAGT 1011
Db 7290 ACTGCTGTCTACTATTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7231
Qy 1012 ACCACTGCTGTCTACTACTTTACCAATTCACCAAGTGTGTATAAACAACCAATCGAA 1071
Db 7230 ATCGCTGCTGTCTACTGCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTACT 7171
Qy 1072 ATTTTGCAACCTATTTCACCAACCACTACCACTTCATATGTTGGTGACTACTTCC 1131
Db 7170 GCTGTCTACTGCTATCACTGCTGTTGCTGCTATCACTGCTGCTGCTGCTGCTATC 7111
Qy 1132 TATCTGACTAAGACTGCGACCAATTTGGTGAACACCTACTGTTATTGTTGATGTCCTAT 1191

```

```
Db 7110 ACTGTGTCACACTGCTATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTATTCCTATTC 7051
Qy 1192 CATACTACCAACACTGTTTACCACTGCAATGGACAGAAACAACTCACTACCAACCAACTGCT 1251
Db 7050 ACTGCTGTCACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6991
Qy 1252 ACCAAATCCAACTGATTCAAATGACACAGTGGTGGTACAAAGTTCCACTGCCAAATCCAAT 1311
Db 6990 GTCACACTGCTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6931
Qy 1312 GTTAGTACTACTGAATATGGTCTCAGTCCCTTTGCTTACCAACCACTACAGTTCCTGCTCT 1371
Db 6930 ACTATTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6871
Qy 1372 CCAGTGTGTACCGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1431
Db 6870 GCTGCTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6811
Qy 1432 GAATATGGTCAAAATCCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1491
Db 6810 GCTGCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6751
Qy 1492 GACTCAGTAATTTATCAGAGAACCAACCAATCCAACTGTCACCTACACACGAGTATGGTCT 1551
Db 6750 GTCACCTGCTATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6691
Qy 1552 CAATCCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1611
Db 6690 ATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6631
Qy 1612 ATCAGAGAACCTCCAAACCACTGTCACCAACCACTGAATATGGTCCCAATCTTACGGA 1671
Db 6630 GCTGCTCACTACTATTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6571
Qy 1672 ACCACAACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1731
Db 6570 GCTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6511
Qy 1732 CCAAAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1791
Db 6510 ACTGCTATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6451
Qy 1792 GTAACCTGACCAACAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1851
Db 6450 GTCACCTGCTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6391
Qy 1852 GTCACCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1911
Db 6390 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6331
Qy 1912 CCAAGTGGCACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1971
Db 6330 ACTATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6271
Qy 1972 GAATACTGCTCAATATGCAACCACTACTACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2031
Db 6270 GTCACACTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6211
Qy 2032 GATACCGTCTTATCAGAGAGCCACCAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2091
Db 6210 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6151
Qy 2092 CAATCATATGCTACAAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2151
Db 6150 ACTGCTGCTACTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6091
Qy 2152 ATCAGAGAGCCCAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2211
Db 6090 GTCACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6031
Qy 2212 ACAACCAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2271
Db 6030 GTCGCTCTGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5971
```

```
Qy 2272 CCAAAATCCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2331
Db 5970 GTCACCTGCTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5911
Qy 2332 GTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2362
Db 5910 GTCACCTGCTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5880

RESULT 13
ABQ73537/c
ID ABQ73537 standard; DNA; 8201 BP.
XX
AC ABQ73537;
XX
DT 03-OCT-2002 (first entry)
XX
DE Human dentin sialophosphoprotein precursor (DSPP) gene SEQ ID NO:1.
XX
KW Human; dentin sialophosphoprotein precursor; dentin sialophosphoprotein;
KW DSPP; dentinogenesis imperfecta type II; deafness; auditory;
KW chromosome 4q21; gene; ds.
XX
OS Homo sapiens.
XX
Key Location/Qualifiers
FH promoter 1..6
FT /*tag= a
FT exon 7..98
FT /*tag= b
FT /*number= 1
FT /*number= 1
FT /*tag= c
FT /*number= 1
FT /*tag= d
FT /*number= 2
FT /*tag= e
FT /*product= "dentin sialophosphoprotein precursor"
FT /*note= "contains introns"
FT sig_peptide 2387..2431
FT /*tag= f
FT mat_peptide 2432..7893
FT /*tag= g
FT /*product= "dentin sialophosphoprotein"
FT /*note= "contains introns"
FT mat_peptide 2432..5520
FT /*tag= h
FT /*product= "dentin sialoprotein"
FT /*note= "contains introns"
FT intron 2438..3576
FT /*tag= i
FT /*number= 2
FT /*tag= j
FT /*number= 3
FT /*tag= k
FT /*number= 3
FT /*tag= l
FT /*number= 4
FT /*tag= m
FT /*number= 4
FT /*tag= n
FT /*number= 5
FT /*tag= o
FT mat_peptide 5521..7893
FT /*product= "phosphophoryn"
FT
```


FT misc_feature 5596..5604
 FT /*tag= p
 FT /note= "cell binding domain"
 FT polyA_signal 7988..7993
 FT /*tag= g
 FT polyA_signal 8171..8176
 FT /*tag= r
 XX
 PN WO200258722-A1.
 XX
 XX 01-AUG-2002.
 PD
 XX
 XX 30-AUG-2001; 2001WO-CN01292.
 PF
 XX
 XX 05-SEP-2000; 2000CN-0125042.
 PR
 XX
 XX (SHAN-) SHANGHAI RES CENT BIOTECHNOLOGY.
 PA
 XX
 XX Kong X, Xiao S, Zhao G, Yu C, Hu L;
 PI
 DR WPI; 2002-557897/59.
 DR P-PSDB; ABP51785.
 XX
 XX
 XX
 PT Diagnosis of dentinogenesis imperfecta type III and its accompanying
 PT deafness using dentin sialophosphoprotein gene and encoded products -
 XX
 XX Example 3; Page 23-27; 38pp; Chinese.
 PS
 XX
 XX The present invention describes a method (M1) for the diagnosis of
 CC dentinogenesis imperfecta type II and/or its accompanying deafness
 CC comprising determining the dentin sialophosphoprotein (DSPP) gene, its
 CC transcript and/or protein of an individual for comparison of their
 CC sequences with the normal sequences and judging the individual to have
 CC higher risk of suffering from the disease then the normal population.
 CC Also described are: (1) treating dentinogenesis imperfecta type III
 CC and/or its accompanying deafness by administering a safe and effective
 CC dose of normal DSPP and/or DSP protein to patients; (2) drug compositions
 CC containing safe doses of DSPP and/or DSP protein; and (3) a reagent kit
 CC for detecting dentinogenesis imperfecta type II and/or its accompanying
 CC deafness containing primers for specific amplification of DSPP gene or
 CC its transcript, or containing probes for binding to the mutation site.
 CC The DSPP gene and protein sequences have auditory activity. The method
 CC (M1), dentin sialophosphoprotein (DSPP) gene and DSP protein are useful
 CC for diagnosing and treating imperfecta type II and/or its accompanying
 CC deafness. The DSPP gene is located to chromosome 4q21. The present
 CC sequence represents the human DSPP gene from the present invention.
 XX
 SQ Sequence 8201 BP; 2801 A; 1477 C; 1873 G; 2050 T; 0 other;
 Query Match 2.5%; Score 94.2; DB 24; Length 8201;
 Best Local Similarity 41.7%; Pred. No. 1.6e-11;
 Matches 588; Conservative 0; Mismatches 823; Indels 0; Gaps 0;
 QY 952 AGTCATGCCGGATCTAACGGTATTGTGCTTCTACAACTAGAACAGTTACAGACAGT 1011
 DB 7290 ACTGCTGTCACTGATGTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7231
 QY 1012 ACCACTGCTGTCACTACTTTACCAATCAATCAAGTGTGATATAAACCAACAAATCGAA 1071
 DB 7230 ATCGCTGCTGTCACTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7171
 QY 1072 ATTTTGGACCTATTCCAACTACCATCACAACTTATATGTTGGTGTGCTACTTCC 1131
 DB 7170 GCTGTCACTGTATCACTGCTGTTGCTGCTATCACTGCTGTCACTGCTGTCACTGCTATC 7111
 QY 1132 TATCTGACTAGACTGACCAATTTGGTGAACAGCTACTGTTATTGTTGATGTGCCATAT 1191
 DB 7110 ACTGCTGTCACTGTATCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTATC 7051
 QY 1192 CATACTACCAACTGTTTACAGTGAATGACAGGAACAATCACTACCAACCAACTCGT 1251
 DB 7050 ACTGCTGTCACTGCTGTTGCTGCTGTCAACCGCTATTGCTGTGCTCACTACTATTACTGCT 6991

QY 1252 ACCAATCCAACTGATTCATTTGACACAGTGGTGTACAGTTCCACTGCCAAATCCAACT 1311
 DB 6990 GTCACCTGCTATCACTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6931
 QY 1312 GTTAGTACTACTGAATATTGGTCTCAGTCTCCTTTGCTACACCCACTACAGTTTACTGCTCT 1371
 DB 6930 ACTATTACTGCTTTCACTGCTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6871
 QY 1372 CCAGGTGGTACCGATACCTGATTTATCAGAGAGCCCAACCACTACTGCTCACTACTACT 1431
 DB 6870 GCTGCTATCACTGCTGTCACTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6811
 QY 1432 GAATATTGGTCAAACTCCTTTGCTTACTACTACTACTACTGCTGCTGCTGCTGCTGCTGCT 1491
 DB 6810 GCTGCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6751
 QY 1492 GACTCAGTAATTTATCAGAGAACCAACCAAACTGCTCACTGCTCACTGCTGCTGCTGCTGCT 1551
 DB 6750 GTCACCTGCTATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6691
 QY 1552 CAATCCCTTTGCTACTACTACTACTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1611
 DB 6690 ATCACTGCTGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6631
 QY 1612 ATCAGAGAACTTCCAAACCAACTGCTCACCACCACTGAATATTGGTCCCCTGCTGCTGCTGCT 1671
 DB 6630 GCTGCTCACTACTATTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6571
 QY 1672 ACCAACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1731
 DB 6570 GCTATCACTGCTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6511
 QY 1732 CCAAAACCACTGCTCACTACTACTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1791
 DB 6510 ACTGCTATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6451
 QY 1792 GTAATCTCACACAGGTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1851
 DB 6450 GTCACCTGCTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6391
 QY 1852 GTCACCTACTACTGATGTTGCTGCTCAATGCTTCTACTACCACTGTAACCTGGTCCA 1911
 DB 6390 GCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6331
 QY 1912 CCAAGTGGCACTGATACTGTTATCATTTAGGAAACCAACCAACCACTGCTCACCACACT 1971
 DB 6330 ACTATCGCTGCTGTTACTATTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6271
 QY 1972 GAATACCTGGTCTCAATCATATGCAACCACTACTACCACTTACCCTGCCCTGGTGAACACT 2031
 DB 6270 GTCACCTACTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6211
 QY 2032 GATACCCGTTCTTATCAGAGAGCCCAACCACTACTGCTCACTACTACTGTAATGCTGCT 2091
 DB 6210 GCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6151
 QY 2092 CAATCATATGCTACACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2151
 DB 6150 ACTGCTGCTCACTACTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6091
 QY 2152 ATCAGAGAGCCCAACCACTACTGCTCACTACTACTGAATGCTGCTCAATCATATGCT 2211
 DB 6090 GTCACCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6031
 QY 2212 ACAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2271
 DB 6030 GTCGCTCTCTGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5971
 QY 2272 CCAATCCACAGTTACTACTGCTGAATATTGGTCACTCAATTTGGCCAAACCAACCACT 2331
 DB 5970 GTCACCTGCTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5911
 QY 2332 GTTACTGCTCCTCCAGGTGGTACTGACACTG 2362

Db 5910 GTCACGCTATCACTGCTGTCACTGTCACTG 5880
|| ||||| | || |||| |||||

RESULT 14

AAA61847
ID AAA61847 standard; DNA; 5511 BP.

XX
AC AAA61847;

XX
DT 28-OCT-2000 (first entry)

XX
DE Cryptosporidium parvum Iowa isolate GP900 ORF.

XX
KW GP900; Iowa isolate; glycoprotein; antibody; cryptosporidiosis;

XX
KW competitive inhibition; attachment; invasion; ligand binding; sporozoite;

XX
KW merozoite; diarrhoea; protozoacide; open reading frame; ORF; ds.

XX
OS Cryptosporidium parvum.

XX
FH Key Location/Qualifiers

XX
FT CDS 1..5511

XX
FT /tag= a

XX
FT /product= "Cryptosporidium parvum Iowa isolate GP900"

XX
FT /note= "No stop codon given"

XX
PN - US6071518-A.

XX
PD 06-JUN-2000.

XX
PF 12-SEP-1997; 97US-0928361.

XX
PR 13-SEP-1996; 96US-0026062.

XX
PR 01-JUN-1993; 93US-0071880.

XX
PR 29-MAY-1992; 92US-0891301.

XX
PR 03-APR-1993; 95US-0415751.

XX
PR 14-AUG-1996; 96US-0700651.

XX
PA (REG) UNIV CALIFORNIA.

XX
PI Petersen C;

XX
XX WPI; 2000-422065/36.

XX
DR P-PSDB; AAB11726.

XX
XX New GP900 protein fragments and fusion proteins of Cryptosporidium

XX
PT parvum, useful for detecting the presence of the parasite, and

XX
PT diagnosing or treating Cryptosporidium infections by competitive

XX
PT inhibition of the function of GP900 -

XX
PS Claim 16; Column 37-42; 59pp; English.

XX
CC The invention relates to the GP900 glycoprotein of the protozoan
CC Cryptosporidium parvum, DNA encoding it, GP900 fragments, and fusion
CC proteins comprising GP900 fragments. The invention also relates to the
CC administration of GP900 or fragments thereof to a host to elicit anti-
CC GP900 antibody production, and to a method of cryptosporidiosis treatment
CC or prophylaxis comprising administration of anti-GP900 antibodies to an
CC individual. Cryptosporidium parvum GP900 and GP900 fragments are able to
CC competitively inhibit sporozoite or merozoite attachment or invasion, and
CC are also useful for the generation of anti-GP900 antibodies. The
CC antibodies also inhibit sporozoite or merozoite attachment/invasion and
CC additionally inhibit the binding of GP900 ligands to GP900. GP900
CC proteins, fragments and antibodies may therefore be used to treat or
CC prevent cryptosporidiosis. Infection with Cryptosporidium is a common
CC cause of diarrhoea in humans and causes life-threatening diarrhoea in
CC immunocompromised persons. Cryptosporidiosis can be contracted from
CC contaminated municipal water supplies (e.g., public swimming pools). It
CC is also a cause of disease in animals, resulting in financial losses in
CC agriculture. GP900 fragments, fusion proteins and antibodies may also be
CC used for the diagnosis of Cryptosporidium parvum infections, and for the
CC detection of the parasite in the environment. The present sequence
CC represents the open erasing frame (ORF) encoding the GP900 protein of the

CC Iowa isolate of Cryptosporidium parvum.

XX
SQ Sequence 5511 BP; 1964 A; 1204 C; 944 G; 1399 T; 0 other;

Query Match 2.3%; Score 87; DB 21; Length 5511;

Best Local Similarity 46.3%; Pred. No. 6.4e-10;

Matches 285; Conservative 0; Mismatches 330; Indels 0; Gaps 0;

QY 986 CTACAACAGAGAGTACAGACAGTACAGACAGTACAGTGTCTGTCTACTACTTTTACCATTCAATCCAA 1045

DB 974 CAACAGGACACACACAACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1033

QY 1046 GTCTTGATAAAACCAAAACAACTGAAATTTTGCACCTATTTCACCAACCACTACTACTACTACTACTACT 1105

DB 1034 CTACTACTAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1093

QY 1106 CTTTCATATGTTGGTGTGCTACTACTTCTCTATCTGACTAGACTGCACCAATTTGGTGAACAG 1165

DB 1094 CCACGACAACTACACCAACCAACCACTACCAAGAACCAACCAACCAACCAACCAACCAACCAACCA 1153

QY 1166 CTACTGTTATTTGTTGATGTGCCATATCATCTACTACCACTGTTTACCAGTGAATGGACAG 1225

DB 1154 CAACT 1213

QY 1226 GAAACAATCTACTACCAACCACTGTTACCAATCCAACTGATTTCAATTTGACACAGTGGTGG 1285

DB 1214 CTACT 1273

QY 1286 TACAAGTTCCACTGCCAAATCCAACTGTTAGTACTACTGTAATTTGGTCTCAGTCTCTTTG 1345

DB 1274 CAACA 1333

QY 1346 CTACAACCACTACTACTGTTCTCCAGGTGTACCGGATCTGTGATTATCAGAGAGC 1405

DB 1334 CAACCAACCACT 1393

QY 1406 CACCAAAACCATACTGTCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1465

DB 1394 CAACA 1453

QY 1466 CTGTTACTGCTCTCCAGGTGTGTTAGTACTACTGACTCAGTAATTTATCAGAGAACCAACCAATCCAA 1525

DB 1454 CCACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACTACT 1513

QY 1526 CTGCTACTACAACCGAGATTTGGTCTCAATCTTTGCTACTACTACTACTACTACTACTACTACTACTCTC 1585

DB 1514 CTACTACTACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACTACTTT 1573

QY 1586 CTCAGGTGGTACTG 1600

DB 1574 CTGAAACTGAGAGTG 1588

RESULT 15

ABT04776

ID ABT04776 standard; DNA; 5511 BP.

XX
AC ABT04776;

XX
DT 27-SEP-2002 (first entry)

XX
DE C parvum GP900 gene fragment SEQ ID NO: 2.

XX
KW Cryptosporidium detection; GP900; P68; cryptopain; cryptosporidiosis;

XX
KW gene; ds.

XX
OS Cryptosporidium parvum.

XX
PN WO200194631-A1.

XX
PD 13-DEC-2001.

XX
PF 14-MAY-2001; 2001WO-US15624.

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2003, 16:36:22 : Search time 90 Seconds
(without alignments)
12900.869 Million cell updates/sec

Title: US-09-715-876-7

Perfect score: 3786

Sequence: 1 atgtctcaacaattacatt.....tgctctgttcatttagtga 3786

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCBUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1069.4	28.2	1071	1	US-08-357-962-1
2	1069.4	28.2	1071	1	US-08-878-106-1
3	348.2	9.2	425	1	US-08-357-962-5
4	348.2	9.2	425	1	US-08-878-106-5
5	87	2.3	5511	3	US-08-928-361B-2
6	87	2.3	7334	3	US-08-928-361B-1
7	86.2	2.3	5163	3	US-08-700-651-1
8	86.2	2.3	5163	3	US-08-928-361B-4
9	86.2	2.3	5318	3	US-08-700-651-2
10	86.2	2.3	5318	3	US-08-928-361B-3
11	85.2	2.3	108	1	US-08-357-962-2
12	85.2	2.3	108	1	US-08-878-106-2
13	82.4	2.2	100	1	US-08-145-705A-32
14	82.4	2.2	100	1	US-08-145-705A-34
15	76	2.0	100	1	US-08-145-705A-33
16	68	1.8	100	1	US-08-145-705A-36
17	58	1.5	3168	4	US-09-165-239A-3
18	57.2	1.5	100	1	US-08-145-705A-37
19	57	1.5	100	1	US-08-145-705A-35
20	54.2	1.4	3489	2	US-08-728-323A-1
21	54.2	1.4	3489	4	US-09-298-568-1
22	54.2	1.4	32207	2	US-08-770-379-20
23	54.2	1.4	32207	4	US-08-757-669A-20
24	54.2	1.4	32207	4	US-09-230-371A-20
25	50.2	1.3	4197	2	US-08-682-517-7
26	50.2	1.3	4197	2	US-08-682-517-8
27	49	1.3	3337	1	US-08-072-610-1

Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 13, Appl
Sequence 83, Appl
Sequence 73, Appl
Sequence 57, Appl
Sequence 103, Appl
Sequence 43, Appl
Sequence 41, Appl
Sequence 141, Appl
Sequence 49, Appl
Sequence 45, Appl
Sequence 59, Appl
Sequence 47, Appl
Sequence 39, Appl
Sequence 51, Appl
Sequence 61, Appl

US-08-719-822B-1
US-09-092-458-1
US-08-911-393-1
US-07-757-022B-13
US-07-757-022B-83
US-07-757-022B-73
US-07-757-022B-57
US-07-757-022B-103
US-07-757-022B-43
US-07-757-022B-41
US-07-757-022B-141
US-07-757-022B-49
US-07-757-022B-45
US-07-757-022B-59
US-07-757-022B-47
US-07-757-022B-39
US-07-757-022B-51
US-07-757-022B-61

ALIGNMENTS

RESULT 1
US-08-357-962-1
; Sequence 1, Application US/08357962
; Patent No. 5668263
; GENERAL INFORMATION:
; APPLICANT: Hoyer, Lois
; APPLICANT: Livi, George
; APPLICANT: Shatzman, Allan
; TITLE OF INVENTION: CONSERVED YEAST NUCLEIC ACID SEQUENCES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/357,962
; FILING DATE: 16-DEC-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jervis, Herbert H
; REGISTRATION NUMBER: 31,171
; REFERENCE/DOCKET NUMBER: P50278
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5019
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1071 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
US-08-357-962-1

Query Match		28.2%	Score 1069.4;	DB 1;	Length 1071;
Best Local Similarity		99.9%;	Pred. No. 5.8e-252;		
Matches 1070;		Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
QY	1300	CCAAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCCTTTGCTACAACTACTACA	1359		
DB	1	CCAAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCCTTTGCTACAACTACTACA	60		
QY	1360	GTTACTGCTCCTCCAGGTGGTACCGGATACGTGATTATCAGAGAGCCACCAACCACTACT	1419		
DB	61	GTTACTGCTCCTCCAGGTGGTACCGGATACGTGATTATCAGAGAGCCACCAACCACTACT	120		
QY	1420	GTCACCTACTACTGAATATTGGTACAAATCCTTTGCTACTACTACTACTGTTACTGCTCCT	1479		
DB	121	GTCACCTACTACTGAATATTGGTACAAATCCTTTGCTACTACTACTACTGTTACTGCTCCT	180		
QY	1480	CCAGGTGGTACTGACTCAGTAATATTATCAGAGAGCCACCAATCCAACTGTCACCTACTACA	1539		
DB	181	CCAGGTGGTACTGACTCAGTAATATTATCAGAGAGCCACCAATCCAACTGTCACCTACTACA	240		
QY	1540	GAGTATTGGTCTCAATCCTTTGCTACTACTACTACTGCTCTCCTCCAGGTGGTACT	1599		
DB	241	GAGTATTGGTCTCAATCCTTTGCTACTACTACTACTGCTCTCCTCCAGGTGGTACT	300		
QY	1600	GACTCAGTAATATTATCAGAGAGCCCTCCAAACCCAACTGTCACCACTGTAATATTGGTCC	1659		
DB	301	GACTCAGTAATATTATCAGAGAGCCCTCCAAACCCAACTGTCACCACTGTAATATTGGTCC	360		
QY	1660	CAATCTTACGCAACCACTACTGCTGCTCTCCAGGAGGCACTGACTCAGTAAT	1719		
DB	361	CAATCTTACGCAACCACTACTGCTGCTCTCCAGGAGGCACTGACTCAGTAAT	420		
QY	1720	ATCAGAGAGCCACCAACCACTGTCACCTACTACTACTGCTGCTCCTCCAGGAGGCACTG	1779		
DB	421	ATCAGAGAGCCACCAACCACTGTCACCTACTACTACTGCTGCTCCTCCAGGAGGCACTG	480		
QY	1780	ACCACTACCACTGTAATGTCACCACTGCTGCTCCTCCAGGAGGCACTGTAATGAGAGCCA	1839		
DB	481	ACCACTACCACTGTAATGTCACCACTGCTGCTCCTCCAGGAGGCACTGTAATGAGAGCCA	540		
QY	1840	CCAAACCACTGTCACCTACTACTGCTGCTCCTCCAGGAGGCACTGTAATGAGAGCCA	1899		
DB	541	CCAAACCACTGTCACCTACTACTGCTGCTCCTCCAGGAGGCACTGTAATGAGAGCCA	600		
QY	1900	GTAACCTGGTCCAACTGTCACCTACTACTGCTGCTCCTCCAGGAGGCACTGTAATGAGAGCCA	1959		
DB	601	GTAACCTGGTCCAACTGTCACCTACTACTGCTGCTCCTCCAGGAGGCACTGTAATGAGAGCCA	660		
QY	1960	GTCACCACTACTGTAATGTCACCACTGCTGCTCCTCCAGGAGGCACTGTAATGAGAGCCA	2019		
DB	661	GTCACCACTACTGTAATGTCACCACTGCTGCTCCTCCAGGAGGCACTGTAATGAGAGCCA	720		
QY	2020	CCTGGTGAACCTGATACCGTCTTATCAGAGAGCCACCAACCACTACTGTCACCTACTACT	2079		
DB	721	CCTGGTGAACCTGATACCGTCTTATCAGAGAGCCACCAACCACTACTGTCACCTACTACT	780		
QY	2080	GAATACCTGGTCTCAATCATATGCTTACCAACCACTGTTACTGCAACCACTGTTGAAACC	2139		
DB	781	GAATACCTGGTCTCAATCATATGCTTACCAACCACTGTTACTGCAACCACTGTTGAAACC	840		
QY	2140	GATACCTGGTCTTATCAGAGAGCCACCAACCACTACTGTCACCTACTACTGTAATGCTCT	2199		
DB	841	GATACCTGGTCTTATCAGAGAGCCACCAACCACTACTGTCACCTACTACTGTAATGCTCT	900		
QY	2200	CAATCATATGCTACCAACCACTGTTACTGCAACCACTGTTACTGCAACCACTGTTATC	2259		
DB	901	CAATCATATGCTACCAACCACTGTTACTGCAACCACTGTTACTGCAACCACTGTTATC	960		
QY	2260	ATTAGAGAGCCACCAATCCAACTGTTACTACTACTACTGTAATATTGGTCAACCACTATTGCC	2319		
DB	961	ATTAGAGAGCCACCAATCCAACTGTTACTACTACTACTGTAATATTGGTCAACCACTATTGCC	1020		
QY	2320	ACAAACCACTGTTACTGCTCCTCCAGGTGGTACTGACACTGTTGATTATC	2370		

Db	1021	ACAACCACCACTGTTACTGCTCCTCCAGGTGGTACTGACACTGTTGATTATC	1071		
RESULT 2					
US-08-878-106-1					
; Sequence 1, Application US/08878106					
; Patent No. 5817466					
; GENERAL INFORMATION:					
; APPLICANT: Hoyer, Lois					
; APPLICANT: Livi, George					
; APPLICANT: Shatzman, Allan					
; TITLE OF INVENTION: CONSERVED YEAST NUCLEIC ACID SEQUENCES					
; NUMBER OF SEQUENCES: 9					
; CORRESPONDENCE ADDRESS:					
; ADDRESSEE: SmithKline Beecham Corporation					
; STREET: 709 Swedeland Road					
; CITY: King of Prussia					
; STATE: PA					
; COUNTRY: USA					
; ZIP: 19406-0939					
; COMPUTER READABLE FORM:					
; MEDIUM TYPE: Diskette					
; COMPUTER: IBM Compatible					
; OPERATING SYSTEM: DOS					
; SOFTWARE: FastSeq Version 1.5					
; CURRENT APPLICATION DATA:					
; APPLICATION NUMBER: US/08/878,106					
; FILING DATE:					
; CLASSIFICATION: 435					
; PRIOR APPLICATION DATA:					
; APPLICATION NUMBER: 08/357,962					
; FILING DATE: 16-DEC-1994					
; ATTORNEY/AGENT INFORMATION:					
; NAME: Jervis, Herbert H					
; REGISTRATION NUMBER: 31,171					
; REFERENCE/DOCKET NUMBER: P50278					
; TELECOMMUNICATION INFORMATION:					
; TELEPHONE: 610-270-5019					
; TELEFAX: 610-270-5090					
; TELEX:					
; INFORMATION FOR SEQ ID NO: 1:					
; SEQUENCE CHARACTERISTICS:					
; LENGTH: 1071 base pairs					
; TYPE: nucleic acid					
; STRANDEDNESS: single					
; TOPOLOGY: linear					
; MOLECULE TYPE: Genomic DNA					
; HYPOTHETICAL: NO					
; ANTI-SENSE: NO					
; FRAGMENT TYPE:					
; ORIGINAL SOURCE:					
US-08-878-106-1					
Query Match					
Best Local Similarity 99.9%;					
Matches 1070; Conservative 0; Mismatches 1; Indels 0;					
QY	1300	CCAAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCCTTTGCTACAACTACTACA	1359		
DB	1	CCAAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCCTTTGCTACAACTACTACA	60		
QY	1360	GTTACTGCTCCTCCAGGTGGTACCGGATACGTGATTATCAGAGAGCCACCAACCACTACT	1419		
DB	61	GTTACTGCTCCTCCAGGTGGTACCGGATACGTGATTATCAGAGAGCCACCAACCACTACT	120		
QY	1420	GTCACCTACTACTGAATATTGGTACAAATCCTTTGCTACTACTACTACTGTTACTGCTCCT	1479		
DB	121	GTCACCTACTACTGAATATTGGTACAAATCCTTTGCTACTACTACTACTGTTACTGCTCCT	180		
QY	1480	CCAGGTGGTACTGACTCAGTAATATTATCAGAGAGCCACCAATCCAACTGTCACCTACTACA	1539		
DB	181	CCAGGTGGTACTGACTCAGTAATATTATCAGAGAGCCACCAATCCAACTGTCACCTACTACA	240		
QY	1540	GAGTATTGGTCTCAATCCTTTGCTACTACTACTACTGCTCTCCTCCAGGTGGTACTGTAAT	1599		
DB	241	GAGTATTGGTCTCAATCCTTTGCTACTACTACTACTGCTCTCCTCCAGGTGGTACTGTAAT	120		
QY	1600	GACTCAGTAATATTATCAGAGAGCCCTCCAAACCCAACTGTCACCACTGTAATATTGGTCC	1659		
DB	301	GACTCAGTAATATTATCAGAGAGCCCTCCAAACCCAACTGTCACCACTGTAATATTGGTCC	60		
QY	1660	CAATCTTACGCAACCACTACTGCTGCTCTCCAGGAGGCACTGACTCAGTAATATTGGTCC	1719		
DB	361	CAATCTTACGCAACCACTACTGCTGCTCTCCAGGAGGCACTGACTCAGTAATATTGGTCC	120		
QY	1720	ATCAGAGAGCCACCAACCACTGTCACCTACTACTACTACTGCTGCTCCTCCAGGAGGCACTG	1779		
DB	421	ATCAGAGAGCCACCAACCACTGTCACCTACTACTACTACTGCTGCTCCTCCAGGAGGCACTG	180		
QY	1780	ACCACTACCACTGTAATGTCACCACTGCTGCTCCTCCAGGAGGCACTGTAATGAGAGCCA	1839		
DB	481	ACCACTACCACTGTAATGTCACCACTGCTGCTCCTCCAGGAGGCACTGTAATGAGAGCCA	60		
QY	1840	CCAAACCACTGTCACCTACTACTGCTGCTCCTCCAGGAGGCACTGTAATGAGAGCCA	1899		
DB	541	CCAAACCACTGTCACCTACTACTGCTGCTCCTCCAGGAGGCACTGTAATGAGAGCCA	60		
QY	1900	GTAACCTGGTCCAACTGTCACCTACTACTGCTGCTCCTCCAGGAGGCACTGTAATGAGAGCCA	1959		
DB	601	GTAACCTGGTCCAACTGTCACCTACTACTGCTGCTCCTCCAGGAGGCACTGTAATGAGAGCCA	60		
QY	1960	GTCACCACTACTGTAATGTCACCACTGCTGCTCCTCCAGGAGGCACTGTAATGAGAGCCA	2019		
DB	661	GTCACCACTACTGTAATGTCACCACTGCTGCTCCTCCAGGAGGCACTGTAATGAGAGCCA	60		
QY	2020	CCTGGTGAACCTGATACCGTCTTATCAGAGAGCCACCAACCACTACTGTCACCTACTACT	2079		
DB	721	CCTGGTGAACCTGATACCGTCTTATCAGAGAGCCACCAACCACTACTGTCACCTACTACT	60		
QY	2080	GAATACCTGGTCTCAATCATATGCTTACCAACCACTGTTACTGCAACCACTGTTATC	2139		
DB	781	GAATACCTGGTCTCAATCATATGCTTACCAACCACTGTTACTGCAACCACTGTTATC	60		
QY	2140	GATACCTGGTCTTATCAGAGAGCCACCAACCACTACTGTCACCTACTACTGTAATGCTCT	2199		
DB	841	GATACCTGGTCTTATCAGAGAGCCACCAACCACTACTGTCACCTACTACTGTAATGCTCT	60		
QY	2200	CAATCTTACGCAACCACTACTGCTGCTCTCCAGGAGGCACTGTAATGAGAGCCA	2259		
DB	901	CAATCTTACGCAACCACTACTGCTGCTCTCCAGGAGGCACTGTAATGAGAGCCA	60		
QY	2260	GTAACCTGGTCCAACTGTCACCTACTACTGCTGCTCCTCCAGGAGGCACTGTAATGAGAGCCA	2319		
DB	961	GTAACCTGGTCCAACTGTCACCTACTACTGCTGCTCCTCCAGGAGGCACTGTAATGAGAGCCA	60		
QY	2320	GTCACCACTACTGTAATGTCACCACTGCTGCTCCTCCAGGAGGCACTGTAATGAGAGCCA	2370		

Qy	1540	GAGTATTGGTCTCAATCCCTTTGCTACTACTACTACAGTTACTGCTCTCCAGGTGGTACT	1599
Db	241	GAGTATTGGTCTCAATCCCTTTGCTACTACTACTACAGTTACTGCTCTCCAGGTGGTACT	300
Qy	1600	GACTCAGTAATTTATCAGAGAAGCTCCAAACCCAACTGTCACCACTCAATATTGGTCC	1659
Db	301	GACTCAGTAATTTATCAGAGAAGCTCCAAACCCAACTGTCACCACTCAATATTGGTCC	360
Qy	1660	CAATCTTAGCGAACCAACAATACTGTGACTGCTCTCCAGGAGGCAGTCACTAGTAATT	1719
Db	361	CAATCTTAGCGAACCAACAATACTGTGACTGCTCTCCAGGAGGCAGTCACTAGTAATT	420
Qy	1720	ATCAGAGAACCCAGCAACCACTGTCACTACTGTAATTAAGTGGTCACAATCATATGCC	1779
Db	421	ATCAGAGAACCCAGCAACCACTGTCACTACTGTAATTAAGTGGTCACAATCATATGCC	480
Qy	1780	ACCACCTACCACTGTAACTGCACCAACAGGTGGTACTGACACTGTTATCATTTAGAGAGCCA	1839
Db	481	ACCACCTACCACTGTAACTGCACCAACAGGTGGTACTGACACTGTTATCATTTAGAGAGCCA	540
Qy	1840	CCAAACCACTGTCACTACTAGTATTGGTCTCAATCGTTTGGTACTTACCACAAC	1899
Db	541	CCAAACCACTGTCACTACTAGTATTGGTCTCAATCGTTTGGTACTTACCACAAC	600
Qy	1900	GTAACTGGTCCACCAAGTGGCACTGATACTGTTATCATTTAGGGAACCAACAAACCAACT	1959
Db	601	GTAACTGGTCCACCAAGTGGCACTGATACTGTTATCATTTAGGGAACCAACAAACCAACT	660
Qy	1960	GTCAACCACTACTGAATTAAGTGGTCTCAATCATATGCAACCACTACTACATTAACGGCTCCA	2019
Db	661	GTCAACCACTACTGAATTAAGTGGTCTCAATCATATGCAACCACTACTACATTAACGGCTCCA	720
Qy	2020	CCTGGTGAACATGATACCGTTCTTATCAGAGAGGCCAACCAACCACTACTGTCACTACTACT	2079
Db	721	CCTGGTGAACATGATACCGTTCTTATCAGAGAGGCCAACCAACCACTACTGTCACTACTACT	780
Qy	2080	GAATACTGGTCTCAATCATATGCTACACCACTGTTACTGCAACCACTGGTGAACCC	2139
Db	781	GAATACTGGTCTCAATCATATGCTACACCACTGTTACTGCAACCACTGGTGAACCC	840
Qy	2140	GATACCGTTCTTATCAGAGAGGCCAACCAACCACTACTGTAATTAAGTGGTCAATATTGGC	2199
Db	841	GATACCGTTCTTATCAGAGAGGCCAACCAACCACTACTGTAATTAAGTGGTCAATATTGGC	900
Qy	2200	CATCATATGCTACAAACCACTGTTACTGCACCAACAGGTGGTACCGGATCTGTTATC	2259
Db	901	CATCATATGCTACAAACCACTGTTACTGCACCAACAGGTGGTACCGGATCTGTTATC	960
Qy	2260	ATTAGAGAGCCACCAATCCACAGTTACTACTGTAATTAAGTGGTCAATCATTTGGC	2319
Db	961	ATTAGAGAGCCACCAATCCACAGTTACTACTGTAATTAAGTGGTCAATCATTTGGC	1020
Qy	2320	ACAACCAACCACTGTTACTGCTCCTCCAGGTGGTACTGACACTGTGATTATC	2370
Db	1021	ACAACCAACCACTGTTACTGCTCCTCCAGGTGGTACTGACACTGTGATTATC	1071

RESULTS

```

US-08-357-962-5
; Sequence 5, Application US/08357962
; Patent No. 5688263
; GENERAL INFORMATION:
; APPLICANT: Hoyer, Lois
; APPLICANT: Livli, George
; APPLICANT: Shatzman, Allan
; TITLE OF INVENTION: - CONSERVED YEAST NUCLEIC ACID SEQUENCES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Smithkline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA

```

COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/357,962
FILING DATE: 16-DEC-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Jervis, Herbert H
REGISTRATION NUMBER: 31,171
REFERENCE/DOCKET NUMBER: P50278
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5019
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 455 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
US-08-357-962-5

Query Match

Query Accn 5.20; Score 340.2; Pred. No. 5.5e-76;
Best Local Similarity 92.8%;

Best local similarity 52.00, Rec.no: 3.50 70,
Matches 376; Conservative 0; Mismatches 28; Indels 1; Gaps 1;

Qy	1980	GTCTCAATCATATATGCAACCACTACTACTACCATTTACCGCTCCACTGGTGAACATGATACCGT	2039
Db	21	GTCACAATCATATGCCCCCACTACCACTGTAACTTCACACCAAGCGTGGTACTGACACTGT	80
Qy	2040	TCCTT-ATCAGAGAGCCACCAACCACTACTGTGCATCTACTGAACTACTGGTCTCAAACTCAT	2098
Db	81	TTATCATTTAGAGAGCCACCAACCACTACTGTCCACCACTACTGAACTACTGGTCTCAGTCTCT	140
Qy	2099	ATGCTACACCACTACTGTTACTTGCACCACTGGTGAACCGATACCGTCTCTTATCAGAG	2158
Db	141	ATGCAACCACTACTACTACATTTACCGCTCCACCTGGTGAACCGATACCGTCTCTTATCAGAG	200
Qy	2159	AGCCACCAACCACTACTGTCTACTACTACTGAACTACTGGTCTCAATCATATGCTACAACCA	2218
Db	201	AGCCACCAACCACTACTGTCTACTACTGAACTACTGGTCTCAATCATATGCTACAACCA	260
Qy	2219	CCACTGTTACTGCACCAACCACTGGTACCAGTACTGTTATCATTTAGAGAGCCACCAAACTC	2278
Db	261	CCACTGTTACTGCACCAACCACTGGTACCAGTACTGTTATCATTTAGAGAGCCACCAAACTC	320
Qy	2279	CAACAGTTACTACTCTGAAATTTGGTCAACAATCAATTTGCCACAACCACTACAGTTACTGT	2338
Db	321	CACAGTTACTACTACTGAATATTGGTCACAACTATTGGCCACCACTACAGTTACTGT	380
Qy	2339	CTCTCCAGTGGTACTGACACTGTGATTTATCTATGAAGCATGT	2383
Db	381	CTCTCCAGTGGTACTGACACTGTGATTTATCTATGAAGCATGT	425

RESULT 4

US-08-878-106-5
; Sequence 5, Application US/08878106
; Patent No. 5817466
; GENERAL INFORMATION:

APPLICANT: Hoyer, Lois
APPLICANT: Livi, George
APPLICANT: Shatzman, Allan
TITLE OF INVENTION: CONSERVED YEAST NUCLEIC ACID SEQUENCES
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Smithkline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/876,106
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/357,962
FILING DATE: 16-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Jervis, Herbert H
REGISTRATION NUMBER: 31,171
REFERENCE/DOCKET NUMBER: P50278
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5019
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 425 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
-TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
US-08-878-106-5

Query Match 9.2%; Score 348.2; DB 1; Length 425;
Best Local Similarity 92.8%; Pred. No. 5.5e-76;
Matches 376; Conservative 0; Mismatches 28; Indels 1; Gaps 1;
QY 1980 GTCTCATATATGCACCACTACTACCATACCGCTCCACCTGGTGCAAACTGATACCGT 2039
Db 21 GTCACATCATATGCCACCACTACCACCTGTAAGTGCACCACTGGTGTACTGACACTGT 80
QY 2040 TCTT-ATCAGAGCCACCAACCACTACTGCTACTACTGTAATCTGCTCTCAATCAT 2098
Db 81 TTATCATATAGAGCCACCAACCACTACTGTCACCACTACTGTAATCTGCTCTCACTCT 140
QY 2099 ATGCTACAACCACTACTGTTACTGCAACCACTGCTGTAACCACTGCTTCTTATCAGAG 2158
Db 141 ATGCAACCACTACTACCATTTACCGCTCCACCTGCTGTAACCACTGCTTCTTATCAGAG 200
QY 2159 AGCCACCAACCACTACTGTCACCTACTACTGTAATCTGCTGTAATCTGCTTATCAGAG 2218
Db 201 AGCCACCAACCACTACTGTCACCTACTACTGTAATCTGCTGTAATCTGCTTATCAGAG 260
QY 2219 CCAGTGTACTGCAACCACTGCTGTAACCACTGCTGTAATCTGCTGTAATCTGCTTATCAGAG 2278
Db 261 CCAGTGTACTGCAACCACTGCTGTAACCACTGCTGTAATCTGCTGTAATCTGCTTATCAGAG 320
QY 2279 CAACAGTACTACTGTAATATTTGGTGCACCAATCATTTGGCACAACCACTGCTGTAATCTG 2338
Db 321 CAACAGTACTACTGTAATATTTGGTGCACCAATCATTTGGCACAACCACTGCTGTAATCTG 380

QY 2339 CTCTCCAGTGGTACTGACACAGTGTGATTATCTATGAAAGCATGT 2383
Db 381 CTCTCCAGTGGTACTGACACAGTGTGATTATCTATGAAAGCATGT 425
RESULT 5
US-08-928-361B-2
Sequence 2, Application US/08928361B
Patent No. 6071518
GENERAL INFORMATION:
APPLICANT: Petersen, Carolyn
TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
TITLE OF INVENTION: SPECIES INFECTIONS
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: PETERS, VERNY, JONES & BIKSA
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-1840
COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928,361B
FILING DATE: 12-SEP-1997
CLASSIFICATION:
PRIOR APPLICATION DATA: US 60/026,062
APPLICATION NUMBER:
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Verny, Hana
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480.76-1(HV)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-1677
TELEFAX: 650-324-1678
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5511 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
-TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-928-361B-2

Query Match 2.3%; Score 87; DB 3; Length 5511;
Best Local Similarity 46.3%; Pred. No. 9e-12;
Matches 285; Conservative 0; Mismatches 330; Indels 0; Gaps 0;
QY 986 CTACACATAGAACAGTTACAGACAGTACCACTGCTGTCTACTACTTTTACCATTCAATCAA 1045
Db 974 CAACAGCAGCAACAAACAACAACTACTACAACTACTACCACTACTACTACTACTACTACT 1033
QY 1046 GTGTGTATAAACCAACAACTGAAATTTTGCACCTATTTCACCACTACTACTACTACTACT 1105
Db 1034 CTACTACTGACAAACCAACCAACAACTACTACTACTACTACTACTACTACTACTACTACT 1093
QY 1106 CTTCATATGTTGTTGCTACTTCTCTATCTGACTAAGACTGCACTACTGTTGGTGAACAG 1165
Db 1094 CCAGCAGCACTACACCAACCAACCACTACTACCAAGAACCAACCAACCAACCAACAA 1153
QY 1166 CTACTGTTATTTGTTGTTGCTGCTATCATCTACTACCACTACTGTTTACCACTGTAATG 1225
Db 1154 CAACACTACTACTACTACCAACCAACCAACCACTACTACTACTACTACTACTACTACTACT 1213
QY 1226 GAACATCACTACTACCACTACTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1285

Query Match 2.3%; Score 86.2; DB 3; Length 5163;
Best Local Similarity 46.0%; Pred. No. 1.4e-11;
Matches 496; Conservative 0; Mismatches 558; Indels 25; Gaps 5;

Qy 968 ACGGTATTGTCATTTGGCTACACTAGACAGTTACAGACAGTACCAGTCTGTCTACTA 1027
Db 189 AAGGTTATAGCATGATTCATCCCGGGTTAGAGTTGTTATCTATCATCTGGTCTCCAA 248
Qy 1028 CTTTACCATTCAATCCCAAGTCTGATATAAACAACAAATCGAAATTTTGCAACCTATTC 1087
Db 249 CTGATCCAT--ACTCTAATTTGCCATTCACCCCTGCTACTGGAAATTTAGTCAGTAGTC 306
Qy 1088 CA--ACCACCTACCATCAACATTCATATGTTGGTGTGACTACTTCCCTATCTGACTAGAAC 1145
Db 307 CACTGGTAAACTATTCCAAACACTTATGCAGGTGTTTATCGTTCTAATGAGACTAAGAC 366
Qy 1146 TGCACCAATTTGGTGAACAGCTACTGTTATGTTGATGTGCCATATCATACTACCAAC 1205
Db 367 CACTGAGCCTAGTCAACACACTTATGCAGGTGTTTATCGTTCTAATGAGACTAAGACCAC 426
Qy 1206 TGTACCAGTGAATGGAGGAGGAACAATCACTACCAACACAACTCGTACCAATCCAATGA 1265
Db 427 TGAGCCTAGTGCACACCAACTTCTGTTGGTAGATCCTAAGATTATGCTCCTGTAA 486
Qy 1266 TTCAATTTGACAGAGTGGTGTACAGTTTCCACTGCCAAATCCAACTGTTAGTACTACTGA 1325
Db 487 TTC---TGAGAACTCTTTTGAACAAGGTCAAAATATTGATGGCAGTAAGTATACAT 543
Qy 1326 ATATTGGTCTCAGTCTTGTGTACAACTACAGTTACTCTCTCTCCAGTGGTACCGA 1385
Db 544 TCCATACATAAATGTTGGAGTGAACACACAACTACCAACAACTACTACTACTAC 603
Qy 1386 TACTGTGATTATCAGAGAGCCCAACCACTACTGTCTACTACTACTGTAATTTGGTCACA 1445
Db 604 TACTACTACGACACACAAACACACGACACACAACTACTACTA-----ACTAC 657
Qy 1446 ATCCTTTGCTACTACTACTGTTACTGTCTCCTCCAGGTGGTACTGACTCAGTAATAT 1505
Db 658 TACCACACTACTACTACGACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 717
Qy 1506 CAGAGAACCAACCACTCACTGCTACCAACCGAGTATTGGTCTCAATCTTTGTACTAC 1565
Db 718 AACACACACACACAGGACTACTACTACTACTACTACTACTACTACTACTACTACTACT 777
Qy 1566 TACTACTACAGTTACTGCTCTCCAGTGGTGTACTGACTCAGTAATTTACAGAACCTCC 1625
Db 778 TACAACCAACAACTACAACCACTACAACCAACAACTACAACCAACCAACCAACCAAC 837
Qy 1626 AAACCCACTGTCAACCACTGCTAGTAAATTTATCAGAGAACCAACCAACCACTACTGT 1685
Db 838 AACCAACAACTACAACCACTACAACCACTACAACCACTACAACCACTACAACCACTACA 897
Qy 1686 GACTGCTCTCCAGGACCTGACTGACTAGTAAATTTATCAGAGAACCAACCAACCACTGT 1745
Db 898 CACACACCACTACCCAGAACCAACCAACCAACCACTACTACACACACACACACAAAC 957
Qy 1746 CACTACTACTGAATGCTGTCACAAATCATATGCCACCCTACTACCAGTGAATGCCACCAC 1805
Db 958 AACTACTACTACAACCAACCAACCAACCAACCACTACAACCACTACAACCAACCAAC 1017
Qy 1806 AGGTGGTACTGACTGTTATCATTTAGAGGCCACCAACCACTACTGCTACTACTACTGA 1865
Db 1018 AACAACTACTACTACTACTACAACCAACCAACCACTACTACTACTACTACAACCAAC 1077
Qy 1866 GTATTGGTCTCAATGTTGTTACTTACCACTGTTACTGGTCCAGCAAGTGCCTACTGA 1925
Db 1078 -----ACTACTACTACTACTACTACAACCAACCAACCACTACTACTACTACTACT 1125
Qy 1926 TACTGTTATCATTTAGGGAACCAACCAACCACTGTCACCACTACTGTAATACTGCTCTCA 1985
Db 1126 AACTACCAGGAACCAACCAACCAACCACTACTACTACTACTACTACTACTACTACTACT 1185

Qy 1986 ATCATATGCAACCACTACTACTACCTATCCGCTCCACCTGGTGAACACTGATACCGTTCTTA 2044
Db 1186 AACAACTACTACTACTGCCACAAACAACTACTACTTCTGAACTGAGAGTGAATTA 1244

RESULT 8

US-08-928-361B-4
Sequence 4, Application US/08928361B.
Patent No. 6071518
GENERAL INFORMATION:
APPLICANT: Petersen, Carolyn
TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
TITLE OF INVENTION: SPECIES INFECTIONS
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: PETERS, VERNY, JONES & BIKSA
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928,361B
FILING DATE: 12-SEP-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Verny, Hana
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480.76-1(HV)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-1677
TELEFAX: 650-324-1678
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 5163 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-928-361B-4

Query Match 2.3%; Score 86.2; DB 3; Length 5163;
Best Local Similarity 46.0%; Pred. No. 1.4e-11;
Matches 496; Conservative 0; Mismatches 558; Indels 25; Gaps 5;

Qy 968 ACGGTATTGTCATTTGGCTACACTAGACAGTTACAGACAGTACCAGTCTGTCTACTA 1027
Db 189 AAGGTTATAGCATGATTCATCCCGGGTTAGAGTTGTTATCTATCATCTGGTCTCCAA 247
Qy 1028 CTTTACCATTCAATCCCAAGTCTGATATAAACAACAAATCGAAATTTTGCAACCTATTC 1087
Db 249 CTGATCCAT--ACTCTAATTTGCCATTCACCCCTGCTACTGGAAATTTAGTCAGTAGTC 305
Qy 1088 CA--ACCACCTACCATCAACATTCATATGTTGGTGTGACTACTTCCCTATCTGACTAAGAC 1145
Db 306 CACTGGTAAACTATTCCAAACACTTATGCAGGTGTTTATCGTTCTAATGAGACTAAGAC 365
Qy 1146 TGCACCAATTTGGTGAACAGCTACTGTTATGTTGATGTGCCATATCATACTACCAAC 1205
Db 366 CACTGAGCCTAGTCAACAACTATATGCAGGTGTTTATCGTTCTAATGAGACTAAGACCAC 425
Qy 1206 TGTTACCAGTGAATGGAGGAGGAACAATCACTACTACCAACCACTCGTACCAATCCAATGA 1265

Db 426 TGAGCCTAGTGCAACACCAACTCTTGTGGTAGATCTTAAGATTAAATGCTCTCTGTAA 485
QY 1266 TTCAATTGACACAGTGGGTGACAGTTCACGTGCCAAATCCAACTGTTAGTACTACTGA 1325
Db 486 TTC---TGAACTCTTTTGAACAGGTCAATATTTGATATGGGAGTAAAGTATACAT 542
QY 1326 ATATTGGTCTAGTCTCTGCTACACCACTACAGTTAGTCTCTCCAGGTGGTACCGA 1385
Db 543 TCCATACACTAAATGTTGGAGTGAACACACAAACAAACAACTACTACTACTAC 602
QY 1386 TACTGTGATATCAGAGAGCCCAACCACTACTGTCTACTACTACTACTACTACTACTACT 1445
Db 603 TACTACTAGACAAACAAACAAACAAACAAACAAACAAACAACTACTACTACTACTACT 656
QY 1446 ATCTTTGGTCTACTACTACTACTCTTACTGCTCTCCAGGTGGTACTGACTCAGTAATTAT 1505
Db 657 TACCCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 716
QY 1506 CAGAGAACCAACCACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1565
Db 717 AACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 776
QY 1566 TACTACTAGAGTTACTGCTCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACTCTC 1625
Db 777 TACAACCAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 836
QY 1626 AAACCAACAACTGTCACCACTACTGTAATTTGGTCCCAATCTTACGCAACCACTACTGT 1685
Db 837 AACCAACAACTTACAACCACTACTACTACTACTACTACTACTACTACTACTACTACTACT 896
QY 1686 GACTGCTCTCCAGGAGGACTGACTCAGTAATTATCAGAGAACCACTACTACTACTACT 1745
Db 897 CACAACCACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 956
QY 1746 CACTACTAGTAATGTTGTCACAACTACTACTACTACTACTACTACTACTACTACTACTACT 1805
Db 957 AACT 1016
QY 1806 AGTGGTACTGACACTGTTATATATTAGAGAGCCCAACCACTACTACTACTACTACTACT 1865
Db 1017 AACAACT 1076
QY 1866 GTATTGGTCTCAATGTTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1925
Db 1077 -----AACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1124
QY 1926 TACTGTTATCATTAGGAAACCAACCACTACTACTACTACTACTACTACTACTACTACTACTACT 1985
Db 1125 AACTACCAAGAAACCAACCACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1184
QY 1986 ATCATATGCAACCACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 2044
Db 1185 AACAACT 1243

RESULT 9

US-08-700-651-2
; Sequence 2, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
; TITLE OF INVENTION: INFECTIONS
; FILE REFERENCE: 480.19-4(HV)
; CURRENT FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415,751
; EARLIER FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 5318
; TYPE: DNA
; ORGANISM: Cryptosporidium parvum
US-08-700-651-2

Query Match 2.3%; Score 86.2; DB 3; Length 5318;
Best Local Similarity 46.0%; Pred. No. 1.4e-11;
Matches 496; Conservative 0; Mismatches 558; Indels 25; Gaps 5;

QY 968 ACGGTATTGTCATTTGCTTACAACTAGAACAGTTACAGACAGTACACACTGCTGTCTACTA 1027
Db 189 AAGGTATTAGCATTTGATTCATCCACGGGTTTAGAGTTGATTCTATCATCTGGTCTCCAA 248
QY 1028 CTTTACCATTCAATCCAAGTGTGATAAAACCAAAACAAATCGAAATTTTCCCACTTATTC 1087
Db 249 CTGATCCAT--ACTCTAATTTGTCATTCACCCCTGTCACCTGGAAATTTAGTCAGTAGGTC 306
QY 1088 CA--ACCCTACCCTACCAACTTCATATGTTGGTGGTACTTCTCTATCTGACTAAGAC 1145
Db 307 CACTGGTAAAACCTATTCCAAACACTTATGCAGGTGTTTATCGTTCCTAATGAGACTAAGAC 366
QY 1146 TGCACCAATTTGGTGAACAGCTACTGTTATTTGTTGATGTCCTATATCATCTACTACCAAC 1205
Db 367 CACTGAGCCTAGTGCACAACTTATGCAGGTGTTTATCGTTCCTAATGAGACTAAGACCAC 426
QY 1206 TGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGTACCAATCCAACTGA 1265
Db 427 TGAGCCTTGTGCAACACCAACTCTTGTGGTAGATCTTAAGATTAAATGCTCTCTGTAA 486
QY 1266 TTCAATTTGACACAGTGGTGTACAAAGTTCACACTGCCAAATCCAACTGTTAGTACTACTGA 1325
Db 487 TTC---TGAACTCTTTTGAACAGGTCAATATTTGATATGGCAGTAGGTATACAT 543
QY 1326 ATATTGGTCTCAGTCTCTTGTCTACAACTACTAGTTACTGCTCTCCAGGTGGTACCGA 1385
Db 544 TCCATACACTAAATGTTGGAGTGAACACACAAACAAACAACTACTACTACTACTACTAC 603
QY 1386 TACTGTGATTATCAGAGAGCCCAACCACTACTGTCACCTACTACTACTACTACTACTACTACT 1445
Db 604 TACTACTGACACAAACAAACAAACAAACAAACAAACAAACAACTACTACTACTACTACTACT 657
QY 1446 ATCTTTGCTACTACTACTACTACTGTTACTGCTCTCCAGGTGGTACTGACTCAGTAATTAT 1505
Db 658 TACCCTACTACTACTAGGACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 717
QY 1506 CAGAGAACCACTAAATCCAACTGTCACCTACTACACCGAGTATTTGGTCTCAATCTTTGCTAC 1565
Db 718 AACAAACAAACAAACAAACGACTACTACTACTACTACTACTACTACTACTACTACTACTACT 777
QY 1566 TACTACTACTAGTTACTGCTCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCTCC 1625
Db 778 TACAACCAACAACTACACCACTACTACACCAACAACTACACCAACAACTACACCAACAAAC 837
QY 1626 AAACCCCACTGTCCACCACTACTGTAATTTGTTGCCAATCTTTACGCAACCACTACTGT 1685
Db 838 AACCAACAACTTACAACCACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 897
QY 1686 GACTGCTCTCCAGGAGGACTGACTCAGTAATTATCAGAGAACCACTACTACTACTACTACT 1745
Db 898 CACAACCACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 957
QY 1746 CACTACTACTGTAATCTGCTGCTACAACTCATATGTCACCACTACTACTACTACTACTACTACT 1805
Db 958 ACT 1017
QY 1806 AGTGGTACTGACACTGTTATCATTTAGAGAGCCCAACCACTACTACTACTACTACTACTACT 1865
Db 1018 AACCACT 1077
QY 1866 GTATTGGTCTCAATGTTTGGTACTTACTACTACTACTACTACTACTACTACTACTACTACTACT 1925

STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/357,962
FILING DATE: 16-DEC-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Jervis, Herbert H
REGISTRATION NUMBER: 31,171
REFERENCE/DOCKET NUMBER: P50278
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5019
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 108 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
US-08-357-962-2

Query Match 2.3%; Score 85.2; DB 1; Length 108;
Best Local Similarity 60.2%; Pred. No. 5.4e-12;
Matches 65; Conservative 41; Mismatches 2; Indels 0; Gaps 0;
QY 1300 CCAATCCAACTGTAGTACTACTGATATTTGGTCTCAGTCTTCTGCTACACCACTACA 1359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 CCAATCMHACMGTYASYACHACYGARTAYTGTCNCARTCTNTWYCHACHACHACHACH 60
QY 1360 GTTACTGCTCTCCAGGTGGTACCGGATCTGCTGATTCAGAGGCA 1407
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 RTDACYGSCWCCWGRGHACYGAYWCHGTDMTYATYAGRGCCW 108

RESULT 12
US-08-878-106-2
Sequence 2, Application US/08878106
Patent No. 5817466
GENERAL INFORMATION:
APPLICANT: Hoyer, Lois
APPLICANT: Livi, George
APPLICANT: Shatzman, Allan
TITLE OF INVENTION: CONSERVED YEAST NUCLEIC ACID SEQUENCES
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ Version 1.5

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/878,106
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/357,962
FILING DATE: 16-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Jervis, Herbert H
REGISTRATION NUMBER: 31,171
REFERENCE/DOCKET NUMBER: P50278
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5019
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 108 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
US-08-878-106-2
Query Match 2.3%; Score 85.2; DB 1; Length 108;
Best Local Similarity 60.2%; Pred. No. 5.4e-12;
Matches 65; Conservative 41; Mismatches 2; Indels 0; Gaps 0;
QY 1300 CCAATCCAACTGTAGTACTACTGATATTTGGTCTCAGTCTTCTGCTACACCACTACA 1359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 CCAATCMHACMGTYASYACHACYGARTAYTGTCNCARTCTNTWYCHACHACHACHACHACH 60
QY 1360 GTTACTGCTCTCCAGGTGGTACCGGATCTGCTGATTCAGAGGCA 1407
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 RTDACYGSCWCCWGRGHACYGAYWCHGTDMTYATYAGRGCCW 108
RESULT 13
US-08-145-705A-32/C
Sequence 3, Application US/08145705A
Patent No. 5489513
GENERAL INFORMATION:
APPLICANT: Springer, Wolfgang; Plempel, Manfred;
APPLICANT: Lberding, Antonius
TITLE OF INVENTION: SPECIFIC GENE PROBES AND
TITLE OF INVENTION: PROCESSES FOR THE DIAGNOSTIC
TITLE OF INVENTION: INVESTIGATION OF CANDIDA
TITLE OF INVENTION: ALBICANS
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: SPRUNG HORN KRAMER & WOODS
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: U.S.A.
ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.4 MB storage
COMPUTER: NEC PowerMate 1 Plus
OPERATING SYSTEM: DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/145,705A
FILING DATE: October 28, 1993
CLASSIFICATION: 535
PRIOR APPLICATION DATA:
APPLICATION NUMBER: German P 42 36 708.5
FILING DATE: October 30, 1992
ATTORNEY/AGENT INFORMATION:

```
; NAME: Kurt G. Briscoe
; REGISTRATION NUMBER: 33,141
; REFERENCE/DOCKET NUMBER: Bayer 8885-KGB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; TELEX:
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Candida albicans
US-08-145-705A-32
Query Match 2.2%; Score 82.4; DB 1; Length 100;
Best Local Similarity 89.0%; Pred. No. 2.5e-11;
Matches 89; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1170 TGTATTGTTGCGATATCATCTACACCAACTGTTACCAAGTGAATGGACAGGAAC 1229
Db 100 TGTATTGTTGATGTTCCATATCATCTACACCAACTGTTACTAGTGAATGGACAGGAAC 41
Qy 1230 AATCACTACCAACCAACTCGTACCAACTCCAACTGATTCA 1269
Db 40 AATCACTACTACTACCAACCAACTATCCAACTGATTCA 1

RESULT 14
US-08-145-705A-34/c
; Sequence 34, Application US/08145705A
; Patent No. 5489513
; GENERAL INFORMATION:
; APPLICANT: Springer, Wolfgang; Plempel, Manfred;
; APPLICANT: L bberding, Antonius
; TITLE OF INVENTION: SPECIFIC GENE PROBES AND
; TITLE OF INVENTION: PROCESSES FOR THE DIAGNOSTIC
; TITLE OF INVENTION: INVESTIGATION OF CANDIDA
; TITLE OF INVENTION: ALBICANS
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: SPRUNG HORN KRAMER & WOODS
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.4 MB storage
; OPERATING SYSTEM: DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/145,705A
; FILING DATE: October 28, 1993
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: German P 42 36 708.5
; FILING DATE: October 30, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurt G. Briscoe
; REGISTRATION NUMBER: 33,141
; REFERENCE/DOCKET NUMBER: Bayer 8885-KGB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; TELEX:
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO

; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Candida albicans
US-08-145-705A-34
Query Match 2.2%; Score 82.4; DB 1; Length 100;
Best Local Similarity 89.0%; Pred. No. 2.5e-11;
Matches 89; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 970 GGTATTGTCATTGTTGCTACAACTAGACACAGTTACAGACAGATACCACTGCTGTCTACTACT 1029
Db 100 GGTATTGTCATTGTTGCTACAACTAGACACAGTTACAGACAGATACCACTGCTGTCTACTACT 41
Qy 1030 TTACCATTCAATCCAAAGTGTGATAAAACCAACAAATCG 1069
Db 40 TTACCATTCAATCCGATGTTGACAAACTAAACAATTG 1
```

```

; ORIGINAL SOURCE:
; ORGANISM: Candida albicans
US-08-145-705A-33

Query Match      2.0%; Score 76; DB 1; Length 100;
Best Local Similarity 85.0%; Pred. No. 9.3e-10;
Matches 85; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Oy 1070 AAATTTTGCAACCTATTCCAACCACTACCATCACAACCTTCATATGTTGGTGTGACTACTT 1129
      ||||||||||| ||||||| ||||||||||| ||||||||||| ||||||||||| |||
Db 100 AAATTTTGCAACCTATTCCAACCACTACCATCACAACCTTCATATGTTGGTGTGACTACTT 1129
      ||||||||||| ||||||| ||||||||||| ||||||||||| ||||||||||| |||

Oy 1130 CCTATCTGACTAAGACTGCACCAATTTGGTGAACAGCTAC 1169
      |||| | |||| ||||| || |||| |||||
Db 40 CCTACAGAACCACCAACTGTACCAATAGGACAAACTGCTAC 1
```

Search completed: April 29, 2003, 19:20:50
Job time : 159 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2003, 19:18:29 ; Search time 259 Seconds
(without alignments)
15906.005 Million.cell updates/sec

Title: US-09-715-876-7

Perfect score: 3786

Sequence: 1 atgcttcacacattacatt.....tgtcttggttcatttagtga 3786

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 709820 seqs, 544064369 residues

Total number of hits satisfying chosen parameters: 1419640

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_NA.*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	78.4	2.1	5403	10 US-09-745-008-33	Sequence 33, Appl
C 2	74.4	2.0	684973	10 US-09-263-959-1	Sequence 1, Appl
C 3	62.2	1.6	2014	10 US-09-842-552-22	Sequence 22, Appl
C 4	55.4	1.5	436	10 US-09-864-761-2885	Sequence 2885, Ap
C 5	54.6	1.4	867	10 US-09-216-393-340	Sequence 340, App
C 6	54.6	1.4	867	10 US-09-216-393-342	Sequence 342, App
C 7	54.6	1.4	1397	10 US-09-216-393-343	Sequence 343, App
C 8	54.6	1.4	1397	10 US-09-216-393-345	Sequence 345, App
C 9	52.8	1.4	4104	10 US-09-801-368-107	Sequence 107, App
C 10	52.2	1.4	1059	12 US-10-073-256-54	Sequence 54, Appl
C 11	52	1.4	2015	10 US-09-842-552-79	Sequence 79, Appl
C 12	50.8	1.3	15720	9 US-10-025-380-1058	Sequence 1058, Ap
C 13	50.8	1.3	15720	10 US-09-922-217-1058	Sequence 1058, Ap
C 14	50.8	1.3	15720	10 US-09-833-263-1058	Sequence 1058, Ap
C 15	50.6	1.3	1236	9 US-10-077-584-3	Sequence 3, Appli
C 16	50.2	1.3	4197	10 US-09-137-531-7	Sequence 7, Appli
C 17	50.2	1.3	4197	10 US-09-137-531-8	Sequence 8, Appli
C 18	48.8	1.3	3183	10 US-09-955-909-1	Sequence 1, Appli
C 19	48.6	1.3	2614	9 US-09-822-846-491	Sequence 491, App

C 20	48.2	1.3	1101	10 US-09-874-062-2	Sequence 2, Appli
C 21	48.2	1.3	6604	10 US-09-880-107-1748	Sequence 1748, Ap
C 22	48	1.3	3331	10 US-09-864-761-19481	Sequence 19481, A
C 23	48	1.3	7104	10 US-09-815-242-4580	Sequence 4580, Ap
C 24	48	1.3	7107	10 US-09-815-242-8291	Sequence 8291, Ap
C 25	47.8	1.3	460	10 US-09-864-761-19383	Sequence 19383, A
C 26	47.8	1.3	2824	12 US-10-124-557-13	Sequence 13, Appl
C 27	47.8	1.3	3066	12 US-10-124-557-83	Sequence 83, Appl
C 28	47.8	1.3	3117	12 US-10-124-557-73	Sequence 73, Appl
C 29	47.8	1.3	3148	12 US-10-124-557-57	Sequence 57, Appl
C 30	47.8	1.3	3420	12 US-10-124-557-103	Sequence 103, App
C 31	47.8	1.3	3813	12 US-10-124-557-43	Sequence 43, Appl
C 32	47.8	1.3	3936	12 US-10-124-557-41	Sequence 41, Appl
C 33	47.8	1.3	3942	12 US-10-124-557-141	Sequence 141, App
C 34	47.8	1.3	3945	12 US-10-124-557-49	Sequence 49, Appl
C 35	47.8	1.3	3963	12 US-10-124-557-45	Sequence 45, Appl
C 36	47.8	1.3	3963	12 US-10-124-557-59	Sequence 59, Appl
C 37	47.8	1.3	4055	12 US-10-124-557-47	Sequence 47, Appl
C 38	47.8	1.3	4086	12 US-10-124-557-39	Sequence 39, Appl
C 39	47.8	1.3	4092	12 US-10-124-557-51	Sequence 51, Appl
C 40	47.8	1.3	4215	12 US-10-124-557-61	Sequence 61, Appl
C 41	47.8	1.3	4575	12 US-10-044-090-303	Sequence 303, App
C 42	47.8	1.3	5008	12 US-10-124-557-1	Sequence 1, Appli
C 43	47	1.2	5361	9 US-09-742-096-2	Sequence 2, Appli
C 44	47	1.2	6060	7 US-08-781-986A-534	Sequence 534, App
C 45	47	1.2	6152	9 US-09-742-096-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-745-008-33/c
; Sequence 33, Application US/09745008
; Patent No. US20020137667A1
; GENERAL INFORMATION:
; APPLICANT: Chuenkova, Marina
; APPLICANT: Pereira, Miercio A.
; TITLE OF INVENTION: T. Cruci-Derived Neurotrophic Agents and
; TITLE OF INVENTION: Methods of Use Therefor
; FILE REFERENCE: 1322.1028-001
; CURRENT APPLICATION NUMBER: US/09/745,008
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/172,881
; PRIOR FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 5403
; TYPE: DNA
; ORGANISM: Trypanosoma cruzi
US-09-745-008-33

Query Match	2.1%	Score 78.4;	DB 10;	Length 5403;
Best Local Similarity	42.2%	Pred. No. 4.2e-07;		
Matches 442;	Conservative 0;	Mismatches 606;	Indels 0;	Gaps 0;
QY 1309	ACTGTTAGTACTACTGAATATTGGTCTCAGTCCTTTGCTACACCACTACAGTTACTGCT 1368			
DB 3789	ACCGTGGGCACCTGCTGTGTCAGGGGAGTCGAGGGGCTACTGTGGGCACCTGCTGTCAGCGGG 3730			
QY 1369	CTCCAGGTGGTACCGGATACCTGTGATTATCAGAGAGAGCCAAACCACTACTGTCTACTACT 1428			
DB 3729	AGTCGAGGGCTACCGTGGGCACCTGCTCAGCGGGAGTCGAGGGGCTACTGTGGGCACCT 3670			
QY 1429	ACTGAATATTGGTGCACAACTCTTTGCTACTACTACTGTTACTGCTCTCTCCAGGTGGT 1488			
DB 3669	GCTGCAACGGGAGTCGAGGGGCTACTGTGGGCACCTGCTGTCAGCGGGAGTCGAGGGGCT 3610			
QY 1489	ACTGACTCAGTATTATCAGAGAACCAACCACTCACTGTCATACCAACGAGTATTGG 1548			
DB 3609	ACCGTGGGCACCTGCTGTCAGCGGGAGTTGAGGGGCTACTGTGGGCACCTGCTGTCACCGGG 3550			

RESULT 4
US-09-864-761-2885
; Sequence 2885, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 2885
; LENGTH: 436
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000018.2
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.4
US-09-864-761-2885

Query Match 1.5%; Score 55.4; DB 10; Length 436;
Best Local Similarity 48.8%; Pred. No. 0.013;

Matches 152; Conservative 0; Mismatches 161; Indels 0; Gaps 0;
QY 965 CTACGGTATTGTCATGTTGCTACAACTAGAACAGTTACAGACAGTACCACCTCTGTCA 1024
DB 68 CTATTGTCATCTATTACCACCTACCATTAGTACTACCACACTATCATCATCTACCATCA 127
QY 1025 CTACTTTACCATTCAATCCCAAGTGTGATATAAAACCAAAACAAATGGAATTTTGCACCTA 1084
DB 128 CTACTAATATCATATTACCACCTACTATTACCATTACCACGACTATCATCTACTACTA 187
QY 1085 TTCCAACCACTACCATCACAACCTTCATATGTTGGTGTGACTACTTCTATCTACTAAGA 1144
DB 188 CCACCTCCACCATTTACTACTAATGCTATCATCTACTACTGCTTCTATAATTA 247
QY 1145 CTGCACCAATTTGGTGAACAGCTACTGTTATGTTGATGTGCCATATCATCTACTACCA 1204
DB 248 CTCCTACTCTACCATTACTACTACTAGTAGTATCACCATTACTACTACTACTACTATTA 307
QY 1205 CTGTTACCAGTGAATGGACAGGACAAATCACTACTACCACCACTCGTACCAATCCA 1264
DB 308 CTATTACCACCACTACTACTATCATCTACTACTACTACCACCACTACTACTACTG 367
QY 1265 ATTCAATTGACAC 1277
DB 368 CTGTCATTACTAC 380
RESULT 5
US-09-216-393-340
; Sequence 340, Application US/09216393
; Patent No. US2001001447A1
; GENERAL INFORMATION:
; APPLICANT: Milhausen, Michael James
; TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND
; FILE REFERENCE: TX-1-C2
; CURRENT APPLICATION NUMBER: US/09/216,393
; CURRENT FILING DATE: 1998-12-18
; EARLIER APPLICATION NUMBER: 08/994,825
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 340
; LENGTH: 867
; TYPE: DNA
; ORGANISM: Toxoplasma gondii
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(867)
US-09-216-393-340
Query Match 1.4%; Score 54.6; DB 10; Length 867;
Best Local Similarity 45.5%; Pred. No. 0.028;
Matches 195; Conservative 0; Mismatches 234; Indels 0; Gaps 0;
QY 3017 CTGATTTGACTAGTGAACCAACAAATACAGAGAACCAACCACTACATTTATCAACTACTT 3076
DB 395 CTGCTTTCATTAGGCCAAGGAGGTCAAAGAGAGGAGGAAAGACTTACAACCTACAACCACT 454
QY 3077 CAACCTCCATCACTGAAGATATACCAACATCTCAACCTACAGGTGATATGAGGACAAATA 3136
DB 455 CATCCACAAGTAGGAGTACCAACGACGACATCAAGTACCCTACTACTACCCTACCACCA 514
QY 3137 CTTTCATCAACCAATCCAGTTCCCAACTGTGGCAACAAGTACTTTAGCATCTGCAAGTGAAG 3196
DB 515 CTAGGACTACTACTACAACTACGACACCAACAACTACAAACCAACCACTACTACAC 574
QY 3197 AAGACAAACAAAGCGTTCTCTGAATCAGCATCCACAAGTTTGAACCAAGTATGGGTG 3256
DB 575 CAACAACAACGACAAACCAACCACTACTACCACCAACCAACCAACCAACCAACCA 634
QY 3257 AAATTCGTGATTAACCTACTCTTCTACTGAAATTTGAGCTACAAACCACTCTACAGAAG 3316

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2003, 17:19:18 ; Search time 3337 Seconds
(without alignments)
18374.622 Million cell updates/sec

Title: US-09-715-876-7

Perfect score: 3786

Sequence: 1 atgttcaacaattacatt.....tgtctgttcattagtga 3786

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	400.6	10.6	959	17 CNS07DAZ	AL440241 T3 end of
c 2	353	9.3	893	17 CNS07D66	AL440068 T7 end of
c 3	272.8	7.2	1050	17 CNS07CMQ	AL439368 T7 end of
c 4	164.6	4.3	1047	17 CNS07D6L	AL440083 T7 end of
c 5	144.4	3.8	1011	17 CNS07DAY	AL440240 T7 end of
c 6	102	2.7	735	17 CNS04NSM	AL299119 Tetraodon

7	100.4	2.7	773	17	CNS01VTG	AL169549 Tetraodon
8	98.4	2.6	908	17	CNS07DJI	AL440548 T7 end of
9	97	2.6	762	17	CNS01XFI	AL171639 Tetraodon
10	94.8	2.5	824	17	AZ185454	AZ185454 SP_1005_A
11	91.4	2.4	501	17	FR0048173	AL444958 Fugu rubr
12	91.2	2.4	989	17	CNS02HAA	AL197365 Tetraodon
13	89.6	2.4	450	17	FR0048073	AL444858 Fugu rubr
14	82.6	2.2	450	17	FR0025683	AL018519 F.rubripe
15	81.4	2.2	641	13	BM181884	BM181884 fv51b11.y
16	76.2	2.0	694	17	BH355163	BH355163 CH230-81P
17	75.6	2.0	530	17	AZ166409	AZ166409 SP_0085_A
18	75.6	2.0	619	17	FR0047601	AL444386 Fugu rubr
19	75	2.0	500	17	B67199	B67199 Cpg0015B Cp
20	74	2.0	935	17	CNS033D4	AL225985 Tetraodon
21	73.4	1.9	619	17	FR0006944	Z90754 F.rubripes
22	72.6	1.9	644	13	BM605179	BM605179 170006870
23	72	1.9	485	17	AZ434703	AZ434703 1M0221C24
24	70.4	1.9	521	17	AZ164800	AZ164800 SP_0075_B
25	69.6	1.8	350	17	BH879665	BH879665 ht47h12.g
26	69.6	1.8	740	10	AV973760	AV973760 AV973760
27	69.2	1.8	788	17	AZ183942	AZ183942 SP_1002_A
28	68.8	1.8	331	17	BH881398	BH881398 hv25c11.b
29	68.4	1.8	933	17	AZ204694	AZ204694 SP_0100_A
30	68	1.8	470	17	FR0018463	AL011359 F.rubripe
31	67	1.8	810	17	AZ193472	AZ193472 SP_1039_B
32	66.4	1.8	358	17	BH777606	BH777606 fzm5013f0
33	65.6	1.7	605	17	AZ640388	AZ640388 1M0502E21
34	65	1.7	642	17	CI1688	AJ226509 Clona Int
35	64.2	1.7	778	13	BM170064	BM170064 EST572587
36	64	1.7	914	17	AZ205202	AZ205202 SP_0100_A
37	63.4	1.7	695	13	BJ388152	BJ388152 B3388152
38	62.4	1.6	273	17	FR0047466	AL444251 Fugu rubr
39	62.4	1.6	619	17	FR0013713	AL004959 F.rubripe
40	62	1.6	415	10	BE556771	BE556771 fk95c07.y
41	61.6	1.6	979	17	CNS06PDW	AL409242 T7 end of
42	61.4	1.6	550	17	FR0043207	AL130899 Fugu rubr
43	61.4	1.6	718	17	FR072907	AZ972907 2M0246M24
44	61.2	1.6	619	17	FR0006038	Z89848 F.rubripes
45	61.2	1.6	832	13	BM170146	BM170146 EST572669

ALIGNMENTS

RESULT 1	CNS07DAZ/c	CNS07DAZ	959 bp	DNA	linear	GSS 08-JUL-2001
LOCUS	T3 end of clone BD0AA010H12 of library BD0AA from strain CBS 94 of					
DEFINITION	Candida tropicalis, genomic survey sequence.					
ACCESSION	AL440241					
VERSION	AL440241.1	GI:12223652				
KEYWORDS	GSS.					
SOURCE	Candida tropicalis.					
ORGANISM	Candida tropicalis					
REFERENCE	1 (bases 1 to 959)					
AUTHORS	Soucié, J.L., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S., de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B., Malpertuy, A., Neuvéglise, C., Ozier-Kalogeropoulos, O., Potier, S., Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M., Wincker, P. and Weissenbach, J.					
TITLE	Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies					
JOURNAL	FEBS Lett. 487 (1), 3-12 (2000)					
MEDLINE	20584711					
REFERENCE	2 (bases 1 to 959)					
AUTHORS	Blandin, G., Ozier-Kalogeropoulos, O., Wincker, P., Artiguenave, F. and Dujon, B.					
TITLE	Genomic exploration of the hemiascomycetous yeasts: 16. Candida tropicalis					

source

misc feature

BASE COUNT	206 a	138 c	290 q	253 t	6 others
BASE COUNT	206 a	138 c	290 q	253 t	6 others

	Best local similarity	60.4%	Freq; no.	3.5e 00,
	Matches	606;	Mismatches	274;
	Conservative	2;	Indels	31;
	Gaps		Gaps	6;

2000 1000 500 0

DD 893 ACATATACCAACACACAGGTGGTACCTGATACCGTGTGTGTCATCGACCCACCAAAATCCA

QY 1525 ACIGTCACCTACAACCGAGTATTGGTCTCAATCCCTTGGCTACTACTACTACAGTTACTTGGCT

DB 833 ACCGTCACIASCCTACIAGAGTATGGT-----TCCGTTGACACTACCATACATATACCAAC

QY 1585 CCTCCAGGTTACCTGACTCAGTAAATTATCAGAGAACCTCCAAACCCCAACTGTTCACCACC

D6 778 CCACCAGGTGGCACTGATACCGTTGTTGTCATCGACCCACCAACCCAACTGTCACCACC

Qy 1645 ACTGAATATTGGTCCCAATCTTACGCACCACTACTGTGACTGCTCCTCCAGGAGGC

D_b 718 ACCGAGTATGGATCTGTTTCTGAACCAACCACCATCACATATACCAACCCACCGTGGT

QY 1705 ACTGACTCAGTAATTATCAGAGAACCACCACTGTCTACTACTGAATACTGG

D**b** 658 ACTGATACCGTTGTTGTCATCGACCCACCAATCCAACCGTNACTACTAGTATGG-

QY 1765 TCACAATCATATGCCACCCTACCACTGTAACCTGCACCACCAGTGGTACTGCACACTGTT

Db 599 - - - - TTCCGTTGAAACTACCATCACATATACCAACCCACCAGGTGGCACTGATACCGTT

QY 1825 ATCATTAGAGGCCACCAACCACTGCTACTACTGAGTATTGGTCTCAATCGTTT

Db 544 GTTGTCATCGACCCACCAATCCAACCGTNACTACTGAGTATGG-----TTCCGTT

Qy 1885 GCTACTACCAACTGTAAGTGGCACTGATCTTATCATTAGGAA

Db 490 GAACTACCATTAACCAACCCAGGTGGTACTGATACCGTTGTTGTCATCGAC

QY .1945 CCACCAACCCAACTGTCAACCACTACTGAATACTGGTCTCAATCATATGCAACCACTACT

OV 2005 ACCATTACCGCTCCACCTGGTGAAC - -TGATACCGTTCTTATCAGACAGCCACCAAC

Db 376 ACATATACCAACCCACCAGGTGGCAGCGTGATACCGTTGTTGTCATCGACCCACCAACC

Qy 2063 ATACTGTCAC TACTACTGAATACTGGTCTCAATCATATGCTACAACCACTGTTACTG

0v 2123 CACCACCCTGCTGAAACCCGATACCCCTTCTTATCAGAGAGCCACCCAAACCAATACCTCTCACTAA

Db 256 ACCCACCAGGTGCACTGATACCGTTGTTGTCATCGACCCACCAACCAACTGTCACCA

Db	75	GATTATGCAACTAGGTTACTGAGACACATGGTCCAGACCAACATGATGTGTGTGT	20
RESULT 4			
CNS07D6L/C			
LOCUS	CNS07D6L	1047 bp	DNA linear
DEFINITION	T7 end of clone BD0AA009H06 of library BD0AA from strain CBS 94 of Candida tropicalis, genomic survey sequence.		
ACCESSION	AL440083		
VERSION	AL440083.1	GI:12223494	
KEYWORDS	GSS.		
SOURCE	Candida tropicalis.		
ORGANISM	Candida tropicalis.		
REFERENCE	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.		
AUTHORS	1 (bases 1 to 1047) Souciet, J.-L., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S., de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B., Malpertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S., Saurin, W., Tekata, F., Toffano-Nioche, C., Wesolowski-Louvel, M., Wincker, P. and Weissbach, J.		
TITLE	Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies		
JOURNAL	FEBS Lett. 487 (1), 3-12 (2000)		
MEDLINE	20584711		
PUBMED	11152876		
REFERENCE	2 (bases 1 to 1047)		
AUTHORS	Blandin, G., Ozier-Kalogeropoulos, O., Wincker, P., Artiguenave, F. and Dujon, B.		
TITLE	Genomic exploration of the hemiascomycetous yeasts: 16. Candida tropicalis		
JOURNAL	FEBS Lett. 487 (1), 91-94 (2000)		
MEDLINE	20584726		
PUBMED	11152891		
REFERENCE	3 (bases 1 to 1047)		
AUTHORS	Genoscope.		
TITLE	Direct Submission		
JOURNAL	Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail: sequef@genoscope.cns.fr - Web: This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.		
COMMENT			
FEATURES	source		
	1..1047		
	/organism="Candida tropicalis"		
	/strain="CBS 94"		
	/db_xref="taxon:5482"		
	/clone="BD0AA009H06"		
	/clone_lib="BD0AA"		
	/note="end : T7"		
	complement(<2..>484)		
	/note="similar to O74660 [Agglutinin-like protein 4 precursor, ALS4] [Candida albicans]"		
	/evidence=not_experimental		
BASE COUNT	373 a 178 c 171 g 322 t		
ORIGIN			
Query Match	4.3%;	Score 164.6;	DB 17; Length 1047;
Best Local Similarity	60.5%;	Pred. No. 2.7e-31;	
Matches	291; Conservative	0; Mismatches 184;	Indels 6; Gaps 1;
QY	2	TGCTTCAACAATTTACATTTGTTATTTCTTATTTGTTCAATTTGCAAGTGCAAGACAATCA	61
Db	483	TTCTCATACAACCCATTGTGTTATTAATGATTCATTTGTTATTTGTTCAAGAGATAT	424

```
QY 62 CTGGTGTGTTTGTAGTCTTTAAATTCATTAACTTGGTCCCAATGCTGCTAATTA-----TC 115
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 423 CTGGGATTTTATGTTTCGATTTCATTAAACATGGAATGCTGCTCGGATTTACCCAGTG 364

QY 116 CTTTCAAGGGCCAGGATACCAACTTGAATGCTGTTTGGTGGTCCCTTAGATGGTA 175
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 363 CATACCAAGTCCACAAATTTCTACTGCACTGCAGAACTAAACATGTTTGAATGGAG 304

QY 176 CCAGTGCCTAATCCAGGGATACATTACATTTGAATATGCTGTTGTTTAAATATACTA 235
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 303 AATCTGCTGAACAGGATGATACATTACATTAATATGCGATGTTATTCAGTTCAATTA 244

QY 236 CTTCAAAATCATCTGTTGATTTAACTCCGATGGTGTGTTAAATATGCTACTGTGCAATTT 295
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 243 CAAATCAAAACACTGTTGATTTGATGCTGATGGCACTACTTATGCGACCTGTAACTCTCA 184

QY 296 ATTCGTGTTGAAGTAATTCACAACTTTTCTACATTACATGCTACTGTGACGAGCTTTGA 355
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 183 ATCTGGGAAGAATTTTACTACTTCTCAAGTTTAAAGTGTACTGTTTCATCTACCTTGA 124

QY 356 AATCATCCATTAAAGGATTTGGTACAGTTTACTTTACCAATTCGATTCATGTTGGTGGAA 415
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 123 CACATATACCCAGCTGCTGGTACTTTTACATGTTTCCCTTGACATTTAATGTTGGTAGTT 64

QY 416 CAGGTTTCATCAACTGATTTGGAAGATTTTAAATGTTTACTGCTGTCACCAATACAGTCA 475
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63 CGGGAACCTCTGTTGTTGACTGATTCCTACTTCTGCTTCAGACAGGTGTAACACAGTAA 4

QY 476 C 476
Db 3 C 3

RESULT 5
CNS07DAY 1011 bp DNA linear GSS 08-JUL-2001
LOCUS T7 end of clone BD0AA010H12 of library BD0AA from strain CBS 94 of
DEFINITION Candida tropicalis, genomic survey sequence.
ACCESSION AL440240
VERSION AL440240.1 GI:12223651
KEYWORDS GSS
SOURCE Candida tropicalis.
ORGANISM Candida tropicalis.
REFERENCE Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
AUTHORS Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)
20584711
PUBMED 11152876
REFERENCE 2 (bases 1 to 1011)
AUTHORS Blandin,G., Ozier-Kalogeropoulos,O., Wincker,P., Artiguenave,F. and
Dujon,B.
TITLE Genomic exploration of the hemiascomycetous yeasts: 16. Candida
tropicalis
JOURNAL FEBS Lett. 487 (1), 91-94 (2000)
MEDLINE 20584726
PUBMED 11152891
REFERENCE 3 (bases 1 to 1011)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
MEDLINE 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
PUBMED seqrefgenoscope.cns.fr - Web :
JOURNAL This GSS is part of a random genomic sequencing program of thirteen
```

yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES
Location/Qualifiers
1..1011
/organism="Candida tropicalis"
/strain="CBS 94"
/db_xref="taxon:5482"
/clone="BD0AA010H12"
/clone_lib="BD0AA"
/note="end : T7"
misc_feature
<652..>988
/note="similar to P46590 [Agglutinin-like protein 1
precursor, ALS1] [Candida albicans]
1 putative frameshift(s)"
/evidence="not_experimental"
BASE COUNT 336 a 188 c 151 g 293 t 43 others
ORIGIN

Query Match 3.8%; Score 144.4; DB 17; Length 1011;
Best Local Similarity 62.6%; Pred. No. 4.5e-26;
Matches 234; Conservative 4; Mismatches 135; Indels 1; Gaps 1;

QY 7 CAACAATTTACATGTTATTCCTATATATTTGTCATATTTGCAATTTGCAAGTCCAAAGACAATCAGTGT 66
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 625 CAAACAACCTCTCTATGCTGTGCATTTGATGCTTACTTCAAGTAAAGAAATTTCCGGT 684

QY 67 GTTTTGTAGTGTAAATTAATTCATTAATGCTCAATGCTGCTCAATGCTTCAAGAGG 126
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 685 GTGTTTCATAGGATTTGAATTCCTTAATCTGGGATAAGAGTGGTAATATGATATCAAGGT 744

QY 127 CCAGGATACCCAACTTTGGAATGCTGTTTGGTGGTTCCTTTAGATGGTACCAGTGCCTCAAT 186
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 745 CCTCAATATCCAAACATGGAATGCCGTGTTGATTTGGTTCGTAGATGSCACAACAATCC 804

QY 187 CCAGGGATACATTCACATGATATGCCATGTCGTTTAAATATATCTACTTCCACAACA 246
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 805 CCAGGTGATACGTTTACTTTCATGTCGCTGCTTCAAGTTCATCTAGTGTGCAAT 864

QY 247 TCTGTTGATTTAACTCCGATGCTGTTAAATATGCTACTTGTCAATTTTAT-TCTGTGA 305
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 865 CTGTTGGATCTGACTGCCAATGKATGACATATGCCACATGTCATCTTCATGCKGGTGA 924

QY 306 AGAATTCACAACATTTTCTACATTAACATGCTACTGTGAACGACGCTTTGAAATCATCCAT 365
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 925 AGAGTTTACTACTTACTCTAGMTTACGTTGCTACTGCTAVAGATTTCGTTAAGTCTGTTTCAT 984

QY 366 TAAGGCATTTGGTA 379
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 985 GAAGCTATGGGATA 998

RESULT 6
CNS04NSM/c 735 bp DNA linear GSS 21-MAY-2000
LOCUS Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION 123M05 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL299119
VERSION AL299119.1 GI:8038260
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis

ACCESSION Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
VERSION Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
KEYWORDS Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
SOURCE Tetraodontidae; Tetraodon.


```

keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES             Location/Qualifiers
     source            1..908
                        /organism="Candida tropicalis"
                        /strain="CBS 94"
                        /db_xref="taxon:5482"
                        /clone="BD0AA013B05"
                        /clone_lib="BD0AA"
                        /note="end : T7"
     misc_feature      <47..>727
                        /note="similar to O74523 [ Agglutinin-like protein 3
                        precursor, Als3 ] [ Candida albicans]"
     evidence=not_experimental
BASE COUNT            264 a    163 c    185 g    288 t      8 others
ORIGIN
Query Match          2.6%; Score 98.4; DB 17; Length 908;
Best Local Similarity 56.0%; Pred. No. 3.7e-14;
Matches 186; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

```

Db	109	AACCTTACAAGAATAGTGCAAATAGTGTGTGATCTGTTGATCTTATGCTCCTGTCGAATCC	168
QY	1956	AACGTGCACCACTACTGGAATACTTGGTCTCAATCATATGCGAACCACTACTACCATTCACGC	2015
Db	169	AACAGTTACGTTGCACAGAGTTTGGGTGGAAGATTTTACAACCTTCGATCACGTTATACAAA	228
QY	2016	TCCACCTGGTGAACACTGATACCCGTTCTTATCAGAGAGCCACCAAAACCATACTGTCACATC	2075
Db	229	TCAACACGCGCAIAGTGATACTGTTATAATCTACGTTCCCCCAAAACCACTGTTACAGT	288
QY	2076	TACTGAATACTGGTCTCAATCATATGCTACAAACCACCACTGTTACTGCACCACTGTGTA	2135
Db	289	CACTGGAAGATGGGTGGGACTTTATGTTAGTGTGCTGATACGCAACTGGAAATTACCTGGAGG	348
QY	2136	AACCGATACCGTTCTTATCAGAGAGCCACCAA	2167
Db	349	AACAGATACAGTAATTCGAATTTGTACCAGAAA	380

RESULT	9
CNS01XF1	
LOCUS	762 bp DNA linear GSS 12-MAY-2000
DEFINITION	Tetraodon nigroviridis genome survey sequence T7 end of clone 202124 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION	AL171639
VERSION	ALI71639.1 GI:7809696
KEYWORDS	GSS: genome survey sequence.
SOURCE	Tetraodon nigroviridis.
ORGANISM	Tetraodon nigroviridis Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE	1 (bases 1 to 762)
AUTHORS	Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 762)

TITLE Saurin, W. and Weissenbach, J.
Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 762)

AUTHORS Roest-Crollius H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C., Bounneau, J., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.

TITLE Characterization and repeat analysis of the compact genome of the

[illegible]

```
Db 29 ACTGAGACNTCAAATAACCTCAGTGGCCTCTAACTCTACTACTAATGCTACTACTACT 88
QY 1336 CAGTCCTTTGCTACAACCACTCAGTACTGCTCTCCAGTGGTACCGACTACTGTGATT 1395
Db 89 ACTGCTACTGCTACTACTACTGCTACTACTACTACTACTACTACTACTACTACTACTACT 148
QY 1396 ATCAGAGAGCCCAACCACTACTGCTACTACTACTACTACTACTACTACTACTACTACT 1455
Db 149 GCTACTACTACTACTACTGCTACTACTACTACTACTACTACTACTACTACTACTACTGCT 208
QY 1456 ACTACTACTACTACTACTGCTCTCCAGTGGTACTGACTCAGTAAATATACAGAGAACCA 1515
Db 209 ACTGCTACTACTACTACTGCTCTCCAGTGGTACTGCTACTACTACTACTACTACTACTACT 268
QY 1516 CCAAAATCCAACTGCTACTCAACCGAGTATTGGTCTCAATCCTTTGCTACTACTACTACA 1575
Db 269 ACTGCTACTGCTGCTACTGCTGCTACTGCTGCTGCTGCTACTACTACTACTACTACTACT 328
QY 1576 GTTACTGCTCTCCAGTGGTACTGCTACTGCTAGTAAATATACAGAGAACCTCCAAACCCCACT 1635
Db 329 GCTACTACTGCTGCTACTACTGCTACTACTACTACTACTACTGCTGCTGCTGCTACTACT 388
QY 1636 GTCAACCACTCAATATTGGTCCCAATCTTAGCGACCACTACTACTGCTACTGCTCTCT 1695
Db 389 GCTACTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 448
QY - 1696 CC 1697
Db 449 NC 450
```

```
RESULT 15
LOCUS BM181884
DEFINITION fv51b11.y1 Sugano SJD adult male danio rerio cDNA clone 5412044 5'
similar to contains element TARI repetitive element ; , mRNA
sequence.
ACCESSION BM181884
VERSION BM181884.1 GI:17512842
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
```

```
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 641)
AUTHORS Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy,
, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wyllie, T., Underwood
, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B.,
Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E.,
Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.
and Wilson, R.
```

```
TITLE WasNU Zebrafish EST Project 1998
JOURNAL Unpublished (1998)
COMMENT Contact: Stephen L. Johnson
Washington University School
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@wustl.edu
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
Sequencing by: Washington University Genome Sequencing Center Clone
distribution information can be found through the I.M.A.G.E.
Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: T3 Et from Amersham
High quality sequence stop: 527.
Location/Qualifiers
1. .641
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="5412044"
/clone_lib="Sugano SJD adult male"
/sex="male"
```

FEATURES

source

```
/tissue_type="whole body"
/dev_host="adult"
/lab_host="DH10B (phage resistant)"
/note="Vector: pME18S-FL3; Site_1: DraIII (CACCATGTG);
Site_2: DraIII (CACTGTGTG); 1st strand cDNA was primed
with an oligo(dT) primer [ATGTGCGCTTTTTTTTTTTTTTTT];
double-stranded cDNA was ligated to a DraIII adaptor
[ATGTGCGCTACTGG], digested and cloned into distinct DraIII
sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
CACCATGTG). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed and donated by Dr. Sumio
Sugano (University of Tokyo Institute of Medical Science).
Custom primers for sequencing: 5' end primer
CTTCTGCTCTAAAGCTGCG and 3' end primer
CGACCTGCGAGCTCGAGCACA."
```

```
BASE COUNT 218 a 234 c 37 g 152 t
ORIGIN
Query Match 2.2%; Score 81.4; DB 13; Length 641;
Best Local Similarity 48.1%; Pred. No. 7.7e-10;
Matches 297; Conservative 0; Mismatches 311; Indels 9; Gaps 2;
QY 2661 ATCCTCAATCGTAACTAGTACTGTTCTTACTGCAAGTACAAATGCTGACTACTCTTCTTC 2720
Db 34 ATCTCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTGC 93
QY 2721 AAGTGATGTTAGTGTCTACATCTTCTGATAATGTTTCAAAATCAGAGATATCAGTTAC 2780
Db 94 TACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTCC 153
QY 2781 AACCGAAACTTCTGTTACAACTATTCAAACTACTTCCAAACCCCAATATATCATCTTCCAGTGAC 2840
Db 154 AACCAACAACACTTCTCCAGCTACTCACTTCTCCAAACAACAACAACACTTCTCCAAACAAC 213
QY 2841 ATCATTCAGTCAGTGTGCTTCAATTCCTCAAGTGTGTTTCAAGAAGTGAAGTAAAGTACAT 2900
Db 214 AACAACTTCTCCAGCTACTCACTCACTTCAACAACAACAACAACAACACT--TCTCCAGCTACATC 270
QY 2901 TACAAGCAATGGAGACAACAACAAGTGTGCTCATGATTCACAATCTACTTCCACTGAAAT 2960
Db 271 AACTTCAACAACAACAACAACACTTCTCCAGCTACAACTTCTCCAGCAACAACAACAACACTTCTCC 330
QY 2961 TGAATTTGTAACAACCACTGTTCTACTAAAGTTTACCACCTGTCGTTTCTTCTTAATACTGA 3020
Db 331 AGCTACATCAACTTCTCCAAACAACAACAACACTTCTCCAGCTACATCAACTTCTCCAAACAAC 390
QY 3021 TTTGACTAGTGAACCAACAATAATACCAGAGACAACAACAACAACAACAACAACAACAACAACA 3080
Db 391 AACAACTTCTCCAGCTACTCACTTCTCCAAACAACAACAACAACAACAACAACAACAACAACA 450
QY 3081 CTCCATCAGTGAAGATATATCACCACATCTCAACCTTACAGGTGATATATGGAGACAATACTTC 3140
Db 451 TTCAACAAC-----AACAAACAACCTTCTCCAGCTACTCAACTTCTCAACAACAACAACAAC 504
QY 3141 ATCAACCAATCCAGTGTCCAACTGTGCAACAAGTACTTTAGCATCTGCAAGTGAAGAAGA 3200
Db 505 TCAGCTACAACTTCTCCAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 564
QY 3201 CAACAAAACGGTTCATGAATCAGCATCCACAAGTTTGAACCAAGTATGGGTGAAGA 3260
Db 565 AACAAACAACCTTCTCCAGCTACTCACTCACTTCAACAACAACAACAACAACAACAACAACAACA 624
QY 3261 TTCTGGATTAACTACTTT 3277
Db 625 TTCTCCAGCAACAACAACAT 641
```

Search completed: April 29, 2003, 20:16:15
Job time : 3377 secs

